

Sparse inverse problems for Fourier imaging: applications to Optical Interferometry and Diffusion Magnetic Resonance Imaging

THÈSE N° 7617 (2017)

PRÉSENTÉE LE 7 AVRIL 2017

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE TRAITEMENT DES SIGNAUX 5
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Anna AURÍA RASCLOSA

acceptée sur proposition du jury:

Prof. D. Van De Ville, président du jury
Prof. J.-Ph. Thiran, Prof. Y. Wiaux, directeurs de thèse
Prof. D. Alexander, rapporteur
Prof. M. E. Davies, rapporteur
Prof. J.-P. Kneib, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

To my dear uncle

Abstract

Many natural images have low intrinsic dimension (a.k.a. *sparse*), meaning that they can be represented with very few coefficients when expressed in an adequate domain. The recent theory of Compressed Sensing exploits this property offering a powerful framework for sparse signal recovery from undetermined linear systems.

In this thesis, we deal with two different applications of remote Fourier sensing, for which the available measurements relate to the Fourier coefficients of our concerned signal: optical interferometry and diffusion Magnetic Resonance Imaging (dMRI). In both applications, we face challenging problems due to a restricted number of available measurements and the nonlinearity of the direct model for the data. Inspired by the Compressed Sensing framework, our strategy to solve these nonlinear and ill-posed problems resorts to reformulating them as linear inverse problems and propose novel priors to leverage the intrinsic low dimensionality of the solution.

The first part of this thesis is devoted to image reconstruction from optical interferometry data. State-of-the-art methods are nonconvex due to the intrinsic data nonlinearity and are therefore known to suffer from a strong sensitivity to initialization. We reformulate the problem as a tensor completion problem, where the aim is to recover a tensor from which we have information through some linear mapping. We propose two different alternatives to solve it, one being a purely convex approach. An original nonconvex alternate minimization method has also been explored. We present results on synthetic data and compare *pros and cons* for both approaches. Our original formulation can be seen as a generalization of the Phase Lift approach and can potentially be applied to other partial phase retrieval problems.

In the second part, we tackle the problem of fiber reconstruction in dMRI. dMRI exploits the anisotropy of the water diffusion in the brain to study the organization of its tissue. Particularly, the goal of our work is to recover the local properties of the axon tracts, i.e. their orientation and microstructural features in every voxel of the brain. We resort to a reweighting scheme to leverage the structured sparsity of the solution, where

the structure originates from the spatial coherence of the fiber characteristics between neighbor voxels. Imposing this original prior promotes a powerful regularization that guarantees a strong robustness to undersampling. Due to a time-consuming measuring process, this ability to solve the imaging problem from few dMRI data points is crucial to guarantee the feasibility of this technique in a clinical context. We present results on real and simulated data and compare our approach to other state-of-the-art methods. We also discuss how our novel approach can actually be applied in a more generic framework for multiple correlated sparse signal recovery.

Keywords: inverse problems, compressed sensing, sparsity, structured sparsity, convex optimization, optical interferometry, diffusion MRI, spherical deconvolution, HARDI, microstructure imaging.

Résumé

Beaucoup d'images naturelles ont une dimension intrinsèque bien réduite (alias *parcimonieuses*), signifiant qu'elles peuvent être représentées avec un nombre limité de coefficients si elles se trouvent exprimées dans une base adéquate. La récente théorie de l'échantillonnage compressif (*compressed sensing*) exploite cette propriété en offrant un cadre très solide pour la récupération des signaux parcimonieux à partir de systèmes linéaires incomplets.

Dans cette thèse, nous traitons deux applications différentes de détection de Fourier à distance, pour lesquelles les mesures qui nous sont accessibles sont liées aux coefficients de Fourier du signal qui nous intéresse : l'interférométrie optique et l'Imagerie par Résonance Magnétique de diffusion (IRMd). Dans les deux cas, nous nous trouvons face à des problèmes qui représentent un gros défi dû à la non-linéarité du modèle et à l'accessibilité à un nombre très restreint de mesures. Inspirés par la théorie de l'échantillonnage compressif, notre stratégie pour résoudre ces problèmes inverses mal posés et non-linéaires recourt à les ré-exprimer comme des problèmes inverses linéaires et à proposer des nouvelles informations à priori afin d'exploiter la petite dimension intrinsèque à la solution.

La première partie de cette thèse est consacrée à la reconstruction des images à partir des données d'interférométrie optique. Dans ce domaine, les méthodes de pointe sont non-convexes dû à la non-linéarité des données et, par conséquent, sont très sensibles à leur initialisation. Nous reformulons ce problème comme un problème de remplissage de tenseurs, avec le but de récupérer un tenseur duquel on obtient information à travers une conversion linéaire. Nous proposons deux alternatives différentes pour le résoudre, dont l'une est purement convexe. Nous avons aussi exploré une nouvelle méthode de minimisation alternée non-convexe. Nous présentons des résultats avec des données synthétiques et on compare les avantages et désavantages des deux approches. Notre formulation originale peut être vue comme une généralisation de l'approche PhaseLift et peut potentiellement être appliquée à d'autres problèmes de récupération de phase.

Dans la deuxième partie, nous abordons le problème de reconstruction des fibres en

IRMd. IRMd exploite l'anisotropie de la diffusion de l'eau dans le cerveau pour étudier l'organisation de ses tissus. En particulier, le but de notre travail est de récupérer les propriétés locales des voies axonales, i.e. leur orientation et microstructure pour chaque voxel du cerveau. Nous recourons à un schéma de repondération itérative afin d'exploiter la parcimonie structurée de la solution, structure qui a lieu grâce à la cohérence spatiale des fibres entre les voxels voisins. En imposant cette nouvelle information a priori nous promouvons une régularisation solide qui assure la reconstruction dans des régimes fortement sous-échantillonnés. À cause d'un processus de mesure de longue durée, le fait d'être capable de résoudre le problème avec très peu de données IRMd est crucial afin de garantir la faisabilité de cette technique dans un contexte clinique. Nous présentons des résultats avec des données autant simulées que réelles que nous comparons à l'état-de-l'art. Nous discutons aussi comment notre nouvelle méthode peut être appliquée à la récupération de multiples signaux parcimonieux corrélés dans un cadre plus générique.

Mots-clefs : problèmes inverses, échantillonnage compressif, parcimonie, parcimonie structurée, optimisation convexe, interférométrie optique, IRM de diffusion, déconvolution sphérique, HARDI, imagerie de la microstructure.

Acknowledgments

Pursuing a PhD is like running a long-distance race. Hereafter, I would like to take the opportunity to thank many people around me who have helped me to get to the finish line since, surely, I would have never gotten there on my own.

First and foremost, I would like to thank my supervisors, Prof. Jean-Philippe Thiran and Prof. Yves Wiaux. Yves, thank you for giving me the opportunity to work in such a privileged international environment, for following closely every detail of my work and always giving valuable scientific advice. Jean-Philippe, thank you for your trust and your unconditional support in all difficult moments throughout my PhD. Thank you for your welcoming smile and your time despite your busy agenda. I would like to acknowledge you as well for making LTS5 a great lab to work in, promoting such a hard-working yet pleasant and fun atmosphere.

I would like to thank the members of the jury, who read through my manuscript, for the stimulating ambiance and discussion during my PhD exam. I thank Prof. Dimitri Van De Ville, president of the jury; Profs. Daniel Alexander, Michael Davies and Jean-Paul Kneib for their insightful questions and comments.

I would like to profoundly thank two very special mentors, Rafael Carrillo and Prof. Alessandro Daducci. You have been my teachers, reviewers and overall good friends throughout these years; without you the road towards my PhD would have been for sure much more challenging and less enjoyable. Thank you, Rafa, for your warm welcome when I first arrived in the lab, for your scientific (and non-scientific) advises and for being such a great office-mate. Thank you, Ale, for always pushing me further, encouraging me and keeping my mood up in difficult moments.

I would like to thank all the other members of LTS5, specially David for our nice discussions and hours of team work together and Marina, my supporting office-mate these last months of my PhD journey (among many other things, thank you for teaching me how to bind a manuscript!). Special thanks as well to always-smiley and helpful Vijay for

your IT support.

Many thanks go to all the present and past members of the LTSs, MMSPG, ASPG and LIONS groups, for our daily interactions and the friendly ambiance in our corridor. A sincere recognition to Rosie, in particular for your huge contribution to keeping the *kitchen* alive, thanks to your lead and energy we all can benefit from a nice break area. A huge thanks to the "kitchen crew" (Ana, Tamara, Meri, Alia, Laura, Elda, Frank,... we used to be many...!) to make lunch breaks entertaining and to the "discofit crew" (we as well used to be many...) to help me stay fit during my PhD. Deep thanks to Eleni, Ale, Fra, Anil, Thomas for your loving company and friendship.

I have only managed to pursue a PhD thanks to resetting my brain from it every day and for that I am profoundly grateful to my dear group of friends in Lausanne. In special, thanks to all Dickens' girls for their open heart, living with you shaped me into who I am. Likewise, I would like to thank my close friends Eva, Sandra and Belén who have supported me despite living further away.

And last but not least, I am deeply thankful to my family. Special thanks to my parents-in-law who helped me out a lot during the writing process and before my private defense, taking care of us. Thank you, Jose and little Enric, for our shared lives and your love. I am profoundly grateful to my uncle (to whom this thesis is dedicated) and aunt, they are my scientific inspiration since I was a little girl and have helped me unconditionally throughout my life, I am not sure I would have ever started a PhD without their influence. And finally, a deep thanks to my parents and siblings, supporting my choice to leave our country and being always, always there despite the distance. Moltes gràcies.

Lausanne, 22 March 2017

Contents

Abstract (English/Français)	v
List of acronyms	xv
List of figures	xviii
1 Preface	1
1.1 Context	1
1.2 Contributions and organization of the manuscript	2
2 Sparse linear inverse problems	5
2.1 Introduction	5
2.2 Notation and basic definitions	5
2.3 Linear inverse problems	8
2.4 Sparse signal model	9
2.4.1 Review on Compressed Sensing	9
2.4.2 Beyond ℓ_1 : iterative reweighted ℓ_1 minimization	11
2.4.3 Low-rank matrix recovery	12
2.5 Convex optimization and proximal splitting methods	13
2.5.1 Forward-Backward algorithm	15
2.5.2 Douglas-Rachford algorithm	16
2.6 Conclusion	16
I Optical Interferometry	19
3 Tensor optimization for optical-interferometric imaging	21
3.1 Introduction	21
3.2 Materials and Methods	24
3.2.1 Data model and tensor formulation	24
3.2.2 Rank-1 alternate minimization (AM)	25

3.2.2.1	Algorithm formulation	25
3.2.2.2	Optimization details	27
3.2.3	Supersymmetric nuclear minimization (NM)	27
3.2.3.1	Algorithm formulation	27
3.2.3.2	Optimization details	30
3.2.3.3	Nuclear minimization with sparsity	31
3.3	Simulations and results	32
3.3.1	AM vs NM comparison	33
3.3.2	NM vs NM-RW comparison	35
3.4	Conclusion	35
II	Diffusion Magnetic Resonance Imaging	41
4	Background on diffusion MRI	43
4.1	Introduction	43
4.2	Principles of diffusion Magnetic Resonance Imaging	43
4.3	dMRI in the brain	45
4.4	Fiber orientation reconstruction	47
4.4.1	Diffusion Tensor Imaging, DTI	47
4.4.2	Diffusion Spectrum Imaging, DSI	48
4.4.3	Methods based on spherical deconvolution	48
4.5	Microstructure imaging techniques	49
4.5.1	The AMICO framework	51
4.6	Conclusion	53
5	Structured sparsity for spatially coherent fiber orientation estimation in diffusion MRI	55
5.1	Introduction	55
5.2	Materials and Methods	57
5.2.1	dMRI framework for recovery of FOD via spherical deconvolution	57
5.2.2	Spatial regularization through structured sparsity	58
5.2.3	Implementation details	62
5.2.4	Phantom data	63
5.2.5	Real Data	63
5.2.6	Evaluation criteria	64
5.2.7	Experimental setup	64
5.3	Results and discussion	65
5.3.1	Phantom data	65
5.3.2	Real Data	68
5.3.2.1	Quantitative comparison	68

5.3.2.2	Qualitative comparison	70
5.4	Conclusion	72
6	Fast microstructure estimation in regions with multiple fibers	75
6.1	Introduction	75
6.2	AMICO _x	76
6.2.1	Materials and Methods	76
6.2.2	Experiments	77
6.2.3	Results and discussion	78
6.3	AMICO _{SAM}	81
6.3.1	Materials and Methods	81
6.3.2	Experiments	84
6.3.3	Results and Discussion	85
6.3.4	Limitations and future work	86
6.4	Conclusion	90
7	Conclusion	93
7.1	Perspectives	94
7.1.1	Optical Interferometry	94
7.1.2	diffusion MRI	95
	Bibliography	107
	Publications	109
	Curriculum Vitae	111

List of Acronyms

2D	2-dimensional
3D	3-dimensional
ADD	Axonal Diameter Distribution
CS	Compressed Sensing
CSF	Cerebrospinal Fluid
DSI	Diffusion Spectrum Imaging
DTI	Diffusion Tensor Imaging
DWI	Diffusion Weighted Imaging
EAP	Ensemble Average Propagator
FOD	Fiber Orientation Distribution
GM	Grey Matter
HARDI	High-Angular Resolution Diffusion Imaging
ODF	Orientation Diffusion Function
OTF	Optical Transfer Function
SD	Spherical Deconvolution
SVD	Singular Value Decomposition
WM	White Matter

List of Figures

2.1	Sparsity of Lena in two different basis	9
2.2	Illustration of the ℓ_p approximation of a point in \mathbb{R}^2 by a one-dimensional space	10
2.3	Sparse signal recovery using a weighted ℓ_1 norm	12
2.4	Projection onto a convex set	14
2.5	Soft-thresholding operator	15
3.1	Sketch of an optical interferometer	22
3.2	Example of variable-density sampling pattern	33
3.3	Reconstruction quality and timing comparison between the NM approach and its analogous considering the unfoldings of the tensor	34
3.4	Reconstruction quality results for synthetic images of size $N = 16^2$	34
3.5	Eta Carinae star system illustration	36
3.6	Rapidly rotating star illustration	37
3.7	M51 Galaxy illustration	38
3.8	Reconstruction quality results for synthetic images of size $N = 16^2$ with randomly distributed spikes	39
3.9	Sparsified version of M51 Galaxy illustration	39
4.1	PGSE sequence	44
4.2	Diffusion of water molecules in the brain	45
4.3	Visualization of the EAP, ODF and FOD functions.	46
4.4	q -space sampling for DWI, DTI and DSI protocols	47
4.5	Schematic representation of the tissue multi-compartment model for ActiveAx [1]	49
4.6	Summary table of the main State-of-the-Art microstructure imaging techniques	52
5.1	Synthetic FOD field in a representative 2D slice	59
5.2	Schematic representation of angular and spatial neighborhoods	60

5.3	Schematic representation of the neighbourhood of element \mathbf{X}_{dv}	61
5.4	Comparison of SR_{\angle} and $\bar{\theta}$ between L2L0, JAMMLSE+L2L0 and L2L0 _{NW} approaches for SNR = 30, 20	67
5.5	Comparison of SR_{\angle} , $\bar{\theta}$, n_{\angle}^{-} and n_{\angle}^{+} between L2L0, JAMMLSE+L2L0 and L2L0 _{NW} approaches for SNR = 10	68
5.6	Comparison of SR_{\angle} and $\bar{\theta}$ between different reconstruction methods	69
5.7	Angular accuracy in real data between L2L0, JAMMLSE+L2L0 and L2L0 _{NW} reconstructions with 30, 20 and 10 samples	70
5.8	Ability of correctly assessing the number of fibers in real data between L2L0, JAMMLSE+L2L0 and L2L0 _{NW} reconstructions with 30, 20 and 10 samples	71
5.9	Qualitative comparison on HARDI human data (30 samples)	72
5.10	Qualitative comparison on HARDI human data (10 samples)	73
6.1	Performance of AMICO ₁ on 2-fiber synthetic substrates	79
6.2	Performance of AMICO _X on 2-fiber synthetic substrates	80
6.3	Robustness to inaccuracies in the estimation of $\boldsymbol{\mu}_2$	80
6.4	Performance of AMICO _X on 3-fiber synthetic substrates	81
6.5	Schematic representation of the recovered coefficients in tensor and matrix form	83
6.6	Schematic representation of a neighborhood in the microstructure space	83
6.7	Comparison between AMICO _X and AMICO _{SAM} for 2 fiber-crossing estimation (270 measurements)	86
6.8	Comparison between AMICO _X and AMICO _{SAM} for 2 fiber-crossing estimation (54 measurements)	87
6.9	Comparison between AMICO _X and AMICO _{SAM} for 2 fiber-crossing estimation (270 measurements)	88
6.10	Comparison between AMICO _X and AMICO _{SAM} for 2 fiber-crossing estimation (54 measurements)	89

1.1 Context

When we have a physical system we can build mathematical models to predict the outcome of a certain measurement process. For instance, the weather forecast predicts the curve of temperatures or the quantity of rain at a certain location based on a particular model. These kinds of problems are commonly known as *forward problems*. Conversely, *inverse problems* are concerned with determining the causes for an observed effect. Inverse problems became very popular by the end of last century being increasingly used for applications in science and engineering. As an example, the restoration of a signal or and image from its "degraded" versions can be cast as an inverse problem and is the basis of many imaging modalities that involve remote sensing, such as medical, astronomical, radar, sonar, etc. In these contexts usually only a limited number of noisy observations is available and this limitation leads to *ill-posed* inverse problems, meaning that their solution either is not unique or does not exists or is not stable. As a consequence, to ensure a plausible solution, the problem needs to be *regularized* by injecting some prior information either on the physical model or on the unknown signal itself.

Many natural signals are *sparse* in some adequate transformed domain, meaning that they can actually be described with very few coefficients when expressed in a suitable basis (i.e. they have *low intrinsic dimension*). For instance, wavelet techniques provide a multi-scale representation of natural signals characterized by a sparse structure. Therefore, functions that promote the sparsity of the signal of interest appear as very good candidates to regularize ill-posed inverse problems.

Among inverse problems in general, linear inverse problems (l.i.p.) have been widely studied and the theory of regularization methods in the linear case is very well developed. One of the nice properties that makes l.i.p very attractive is that they can be efficiently solved by using fast and versatile convex optimization methods. The theory of Compressed Sensing (CS)* can actually be framed in the context of ill-posed linear sensing. CS builds a framework where a sparse signal can be recovered from very few measurements, beyond

*Also known as *compressive sensing* or *compressed/ive sampling*. Throughout this manuscript the reader will find these terms used indistinctly.

the requirements of the Nyquist sampling theorem, by exploiting the sparsity of the signal and cleverly designing adequate sensing matrices.

In this thesis, we deal with sparse, ill-posed and nonlinear inverse problems that arise from two applications of Fourier sensing: optical interferometry and diffusion Magnetic Resonance Imaging (dMRI). In what we call Fourier sensing, not directly the signal of interest is retrieved but a set of measurements that are related to the Fourier coefficients of the object of interest. In both applications, we deal with *undersampled* problems, for which the number of available measurements is smaller than the dimension of the signal. Nevertheless, the nature of the ill-posedness stems from different origins, depending on the application. In optical interferometry the undersampling is due to an highly-limited number of sensing devices (i.e. telescopes). Otherwise, in the field of diffusion MRI the measuring process is so time consuming that undersampling is the only way to guarantee the feasibility of the technique in a clinical context.

We tackle these originally nonlinear inverse problems by *reformulating them as linear problems*, so that we can take advantage of all the flexibility of convex optimization methods to easily inject prior information and efficiently solve them. Our formulations to solve sparse linear inverse problems are built on the CS framework, even if in this work we do not explore explicitly the design of sensing matrices. The strategies that have been implemented to move from a nonlinear to a linear formulation, as well as the chosen regularization methods to leverage the appropriate sparse structure of the signal, vary depending on the application. The reader can find an overview of all of them in the next subsection, devoted to describe our contributions and the organization of the manuscript.

1.2 Contributions and organization of the manuscript

In the following lines I present a brief summary of each of the parts of this manuscript:

Chapter 2 includes a review on general linear inverse problems and describes the theory of Compressed Sensing (CS) for sparse signals. From the formulation perspective, it addresses the problem of sparse signal recovery based on CS ideas. Regarding the reconstruction, it revises convex optimization and proximal splitting methods. This part is meant to provide the reader with the background required to understand subsequent parts of the thesis and does not include any original contribution.

In **part I**, I detail the work we have done on *image reconstruction from Optical Interferometry data*. Optical interferometers are now the state-of-the-art to reach the best resolution in the optical spectrum. Nevertheless, their output is not directly an image and tools for image reconstruction are, therefore, required. Image recovery in optical interferometry is an ill-posed problem arising from incomplete power spectrum and bispectrum

measurements. Consequently, the direct model of the data happens to be highly nonlinear, inducing nonconvexity in the problem. So far, all state-of-the-art methods perform local optimization, which (i) does not provide a global solution and (ii) makes them highly dependent on the initialization. We generalize the Phase Lift approach [2] and reformulate the problem as a *tensor completion problem*, where the aim is to recover a tensor from which we have information through some linear mapping. In order to regularize it, we propose a novel prior to account for low-rankness of the tensor, trying to exploit the specific conditions of our problem. For the sake of comparison, an original nonconvex gauss-seidel alternate minimization approach has also been explored. We present results on synthetic data and discuss *pros and cons* of both approaches. This work has been published in [3, 4].

In **part II**, I describe the work we have done on *fiber reconstruction from diffusion Magnetic Resonance Imaging (dMRI) data*, in the domain of brain imaging. dMRI exploits the random diffusional property (Brownian motion) of the water molecules to try to infer features of the body tissue. In particular, it has become popular in brain imaging to recover axons orientation and microstructure. In **chapter 4** we give some background information on the field and cite the main state-of-the-art local modeling methods to explain the dMRI signal in each voxel of the brain.

Chapter 5 addresses the problem of fiber orientation estimation. The direct forward model that explains the dMRI signal as a Gaussian mixture is actually nonlinear and estimating its parameters directly resorts to nonlinear fitting. However, the problem can be discretized and reformulated using a linear dictionary made of a concatenation of "response functions" in the framework of spherical deconvolution [5, 6]. In this framework, we take advantage of the versatility of convex optimization to include a *novel prior* that imposes spatial regularization directly on the fiber space. We present results on real and simulated data and compare our approach to other state-of-the-art methods. This work has been published in [7, 8].

Chapter 6 is concerned with assessing the microstructure of the fibers. Most microstructure imaging techniques recover the microstructure properties by modeling the dMRI signal in different tissue compartments. Similarly as in chapter 5, the direct nonlinear routines usually employed to fit these models are computationally very intensive and cause practical problems for their application in clinical studies. Our novel approach is based on the framework of AMICO[†] that reformulates these microstructure imaging techniques as linear systems [9]. Again, thanks to the flexibility of convex optimization, we generalize its formulation to enable microstructure estimation and fiber orientation recovery simultaneously. Besides, we adapt the novel prior defined in chapter 5 to the problem of microstructure recovery to impose spatial coherence on the solution. We show preliminary results on simulated data and discuss general open questions in the field of microstructure imaging. Part of this work has been published in [10, 11, 12, 13].

[†]Accelerated Microstructure Imaging via Convex Optimization

Finally, conclusions and future perspectives are presented in **Chapter 7**.

Sparse linear inverse problems

2.1 Introduction

Many natural signals have low intrinsic dimension (a.k.a. sparse), meaning that most of their coefficients vanish when expressed in a suitable basis. As a consequence, the concept of *sparsity* has become very popular in many signal processing problems, such as deconvolution, denoising and deblurring. Sparsity of the signal can appear in different domains: besides classical sparsity, TV-sparse signals (i.e. signals whose gradient has few significant coefficients) or low-rank matrices can also be considered sparse objects since they have much less degrees of freedom than their actual dimension.

Sparsity is therefore a good candidate to be injected as a prior to regularize ill-posed inverse problems. Precisely, the recent theory of Compressive Sensing (CS) resorts to ℓ_1 -norm minimization to promote sparse solutions. In this chapter we address the problem of sparse signal recovery from both, formulation and reconstruction perspective.

The chapter is organized as follows: our notation and some useful mathematical definitions are presented in section 2.2. In section 2.3 the reader can find a short reminder on linear inverse problems and regularization strategies. Section 2.4 is devoted to the formulation of convex minimization problems for the recovery of sparse signals based on Compressive Sensing ideas. Low-rank matrix recovery is also addressed, seen as an extension of the original CS framework. In section 2.5 we revise fundamental notions on convex optimization. We present, as well, the group of *proximal splitting methods* that appear as reconstruction algorithms particularly convenient to solve the convex problems previously proposed. We conclude in section 2.6 reviewing the background context presented in this chapter and linking it with the novel contributions that will be introduced further in this thesis.

2.2 Notation and basic definitions

Throughout this thesis, we use \mathbb{R} to denote the set of real numbers and \mathbb{C} for the complex set. \mathbb{R}_+ denotes the set of non-negative real numbers. We denote **vectors** with bold lower case letters, matrices with upper case letters and tensor with italic ones. We denote

x_i the i -th element of a vector $\mathbf{x} \in \mathbb{R}^N$ (or alternatively \mathbb{C}^N) for any index $i \in \{1, \dots, N\}$. Let $\mathbf{x} \in \mathbb{R}^N, \mathbf{z} \in \mathbb{C}^N$ be two N -dimensional real and complex vectors, respectively. \mathbf{x}^\top stands for the simple transpose vector whereas \mathbf{z}^\dagger stands for the conjugate-transpose vector.

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ be two N -dimensional vectors. The standard inner product on \mathbb{R}^N is given by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$. The ℓ_2 norm, the Euclidean norm, of a vector $\mathbf{x} \in \mathbb{R}^N$ is defined as $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\top \mathbf{x}} = \sqrt{\sum_{i=1}^N x_i^2}$. More generally, the ℓ_p norm ($p \geq 1$) of a vector is defined as:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^N |x_i|^p \right)^{1/p}. \quad (2.1)$$

In particular, for the case $p = 1$, the well-known ℓ_1 norm corresponds to the sum of the absolute values of the signal. The ℓ_0 pseudo-norm of a vector is a cardinality function which corresponds to the number of its nonzero coefficients:

$$\|\mathbf{x}\|_0 = \#(i | x_i \neq 0). \quad (2.2)$$

For **matrices** $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{M \times N}$, the standard inner product is given by:

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{X}^\top \mathbf{Y}) = \sum_{i=1}^M \sum_{j=1}^N \mathbf{X}_{ij} \mathbf{Y}_{ij}, \quad (2.3)$$

where tr denotes the trace of a matrix (i.e, the sum of the elements of its diagonal). \mathbf{X} can be factored as

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top, \quad (2.4)$$

where $\mathbf{U} \in \mathbb{R}^{M \times r}$ satisfies $\mathbf{U} \mathbf{U}^\top = \mathbf{I}$ (being \mathbf{I} the identity matrix), $\mathbf{V} \in \mathbb{R}^{N \times r}$ satisfies $\mathbf{V} \mathbf{V}^\top = \mathbf{I}$ and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r)$ with

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0. \quad (2.5)$$

Factorization (2.4) is known as the *singular value decomposition* (SVD) of \mathbf{X} and σ_i are the singular values. The rank r of a matrix denotes the number of its nonzero singular values. The nuclear norm of a matrix $\|\cdot\|_*$ is equal to the sum of its singular values:

$$\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X}), \quad (2.6)$$

for a matrix \mathbf{X} of $\text{rank}(\mathbf{X}) = r$. Note that the rank of a matrix actually corresponds to the ℓ_0 norm of its vector of singular values, whereas the nuclear norm of a matrix is equivalent

to the ℓ_1 norm of its vector of singular values.

A symmetric matrix $\mathbf{Z} \in \mathbb{R}^{N \times N}$ (or more generally $\in \mathbb{C}^{N \times N}$) is called *positive definite* if for all $\mathbf{x} \neq 0$, $\mathbf{x}\mathbf{Z}\mathbf{x}^\top \geq 0$ (or $\mathbf{z}\mathbf{Z}\mathbf{z}^\dagger \geq 0$) and we denote $\mathbf{Z} \succ 0$. The *negative definite*, *positive semidefinite*, and *negative semidefinite* matrices are defined analogously, requiring the expression $\mathbf{z}\mathbf{Z}\mathbf{z}^\dagger$ to be always negative, nonnegative, and nonpositive, respectively.

The order (or number of dimensions, ways or modes) of a **tensor** $\mathcal{X} \in \mathbb{C}^{N_1 \times \dots \times N_d}$ with components $\mathcal{X}_{i_1, \dots, i_d}$ is the number d of the indices characterizing its components. For the sake of simplicity, we present the following notation only for tensors of order 3.

A 3-way tensor $\mathcal{X} \in \mathbb{C}^{N_1 \times N_2 \times N_3}$ is rank-1 if it can be written as the outer product of 3 vectors, i.e. $\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$, or component-wise $\mathcal{X}_{ijk} = a_i b_j c_k$.

The rank of a tensor, $\text{rank}(\mathcal{X})$, is defined as the smallest number of rank-1 tensors that generate \mathcal{X} as their sum. In other words, if \mathcal{X} can be expressed as

$$\mathcal{X} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (2.7)$$

then $\text{rank}(\mathcal{X}) \leq R$. The notion of rank when applied to a tensor is analogue to the matrix rank though most of the common properties of the latter do not hold when dealing with objects of a dimension higher than 2. One of the main differences is that there is no algorithm to compute the rank of a given tensor. In fact the problem is NP-hard [14]. The well-known method to find a rank- k approximation of a matrix through the largest k values of its SVD [15] does not apply or have an equivalent for the case of high-dimension tensors.

Matricization is the process of transforming a tensor into a matrix. The mode- n matricization of a tensor \mathcal{X} is denoted by $\mathbf{X}_{(n)}$ and results from unfolding all its modes but the mode n into the rows of a matrix. The n -rank of a tensor follows as

$$n\text{-rank}(\mathcal{X}) = (\text{rank}(\mathbf{X}_{(1)}), \text{rank}(\mathbf{X}_{(2)}), \text{rank}(\mathbf{X}_{(3)})). \quad (2.8)$$

In contrast to the rank function, it is easier to handle, since the problem is reduced to calculations with matrices which are already well-known objects with nice properties. The reader can refer to [16] for a more detailed explanation on different notions of tensor rank and their associated decomposition methods.

A tensor is called cubical if every mode has the same size, i.e. $\mathcal{X} \in \mathbb{R}^{N \times N \times N}$. A cubical tensor \mathcal{X} is called supersymmetric if its entries are invariant under permutation of their indices: $\mathcal{X}_{ijk} = \mathcal{X}_{ikj} = \mathcal{X}_{jik} = \mathcal{X}_{jki} = \mathcal{X}_{kij} = \mathcal{X}_{kji}$.

2.3 Linear inverse problems

A typical linear inverse problem aims to solve a linear system of equations of the form

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\eta}, \quad (2.9)$$

where $\Phi \in \mathbb{R}^{M \times N}$ and $\mathbf{y} \in \mathbb{R}^M$ are known, $\boldsymbol{\eta}$ is some noise or perturbation of the measurement vector \mathbf{y} and $\mathbf{x} \in \mathbb{R}^N$ represents the true signal to be recovered. The least squares (LS) approach to solve problem (2.9) chooses to minimize the data discrepancy by solving

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\Phi \mathbf{x} - \mathbf{y}\|^2. \quad (2.10)$$

However, in many applications, matrix Φ is ill-conditioned, implying that small changes in the input argument \mathbf{x} can result in large changes on the output value of the measurements, making (2.9) a very unstable system, extremely sensitive to noise.

Regularization methods overcome this challenge by replacing the original ill-posed problem by a well-conditioned problem with solution

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\Phi \mathbf{x} - \mathbf{y}\|^2 + \lambda f_r(\mathbf{x}). \quad (2.11)$$

In (2.11) the role of the regularization function $f_r(\cdot)$ is to improve the ill-conditioned nature of the original problem by injecting some **prior knowledge** on the required solution. The so-called regularization parameter λ controls the compromise between *accuracy* and *stability* of the solution. Actually, the whole "art" of regularizing relies on wisely tuning this compromise [17]. To mention a simple example, one popular strategy, known as Tikhonov regularization, corresponds to choosing $f_r(\cdot) = \|\cdot\|_2^2$, the quadratic penalty of the unknown:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\Phi \mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathbf{x}\|_2^2. \quad (2.12)$$

In (2.12), the quadratic term that is added to the initial objective function helps stabilizing the problem by controlling the norm of the solution. In this case, λ can be read off as a trade-off between data fidelity and noise sensitivity. Problem (2.11) has also an equivalent constrained form that reads:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} f_r(\mathbf{x}) \quad \text{s.t.} \quad \|\Phi \mathbf{x} - \mathbf{y}\|^2 \leq \epsilon, \quad (2.13)$$

for some value of ϵ acting as a bound on the noise level.

In the following section, we focus on the design of minimization problems, i.e. the choice of relevant regularization functions, for sparse signal recovery.

2.4 Sparse signal model

2.4.1 Review on Compressed Sensing

It is known that a large variety of natural signals are sparse or compressible in multi-scale dictionaries, such as wavelet bases. By definition, a signal is *sparse* in some orthonormal basis $\Psi \in \mathbb{C}^{N \times N}$ if its expansion $\alpha \in \mathbb{C}^N$, with $x \equiv \Psi\alpha$, contains only a small number $K \ll N$ of nonzero coefficients α_i . More generally, it is *compressible* if its expansion only contains a small number of significant coefficients, i.e. if a large number of its coefficients bear a negligible value. In figure 2.1, we show an example of an image and its sparse representations in two different bases.

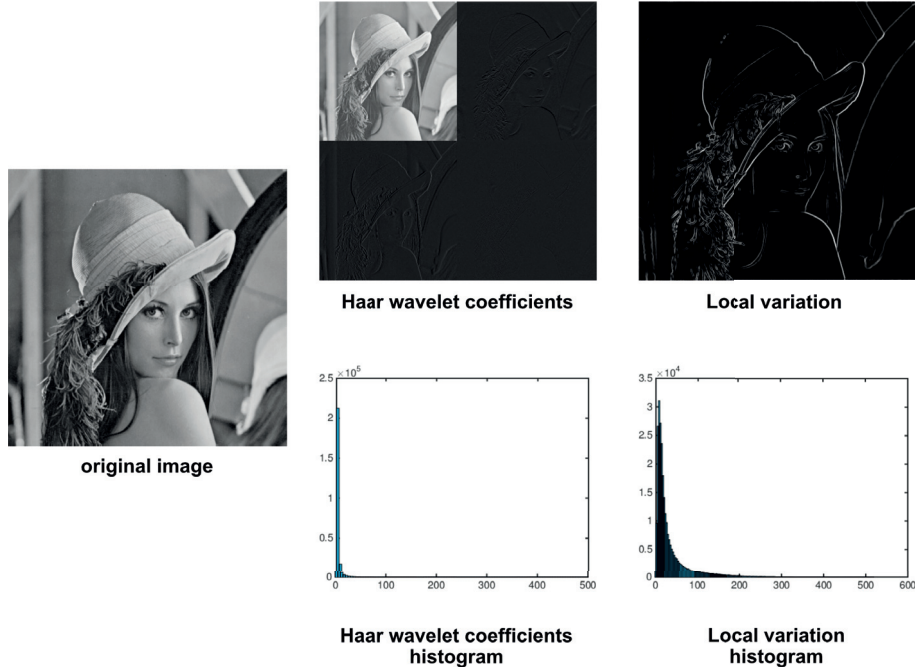


Figure 2.1: Sparsity of Lena in two different basis

The Compressed Sensing (CS) theory [18, 19, 20] builds a framework where a signal can be recovered with very high probability from fewer measurements $M \propto K \ll N$ than what it had used to be considered sufficient, beyond the traditional Nyquist paradigm [21]. The signal is assumed to be probed by M linear measurements denoted by a vector $y \in \mathbb{C}^M$ in some orthonormal sensing basis and identified by the sensing matrix $\Phi \in \mathbb{C}^{N \times M}$ and possibly affected by i.i.d. Gaussian noise $\eta \in \mathbb{C}^M$. Therefore the following linear inverse problem in matrix form holds:

$$y \equiv \Theta\alpha + \eta, \quad \text{with} \quad \Theta \equiv \Phi\Psi \in \mathbb{C}^{M \times N} \quad (2.14)$$

One popular strategy for the signal recovery resorts to solving the associated minimization problem and regularizing the originally ill-posed inverse problem by an explicit **sparsity or compressibility prior** on the signal. Knowing that the representation of signal α is sparse or compressible and considering a set of measurements \mathbf{y} corrupted by some bounded noise, the first natural approach to recover \mathbf{x} would be by solving an optimization problem like:

$$\min_{\hat{\alpha} \in \mathbb{C}^N} \|\hat{\alpha}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \Theta \hat{\alpha}\|_2 \leq \epsilon, \quad (2.15)$$

where ϵ represents a bound on the ℓ_2 norm of the residual noise, $\boldsymbol{\eta} \equiv \mathbf{y} - \Theta \hat{\alpha}$. Unfortunately, the ℓ_0 norm is a nonconvex function and it is known that finding a minimum for (2.15) is NP-hard.

One common approach to recover \mathbf{x} from \mathbf{y} is to replace the nonconvex $f_r(\cdot) = \|\cdot\|_0$ in (2.15) by its convex approximation, the ℓ_1 norm, and solve the following convex problem:

$$\min_{\hat{\alpha} \in \mathbb{C}^N} \|\hat{\alpha}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Theta \hat{\alpha}\|_2 \leq \epsilon. \quad (2.16)$$

In (2.16), the ℓ_1 norm of the coefficients of the signal in the sparsity basis $\alpha \equiv \Psi^\dagger \mathbf{x}$ is minimized under a constraint on the ℓ_2 norm of the residual noise. This ℓ_1 minimization approach formulates the CS problem of recovering sparse signals in the powerful framework of convex optimization, for which there exists a number of efficient numerical solvers and algorithms (refer to section 2.5). A simple illustrated example meant to provide an understanding on how the ℓ_1 norm induces sparsity on the solution can be found in Figure 2.2. In the illustration, dashed lines represent an ℓ_1 ball and an ℓ_2 ball (set of vectors with the same ℓ_1 or ℓ_2 norm, respectively) and line A represents a linear data constraint. The constrained line intersects the ℓ_1 ball on the axis, thus resulting in a sparse solution.

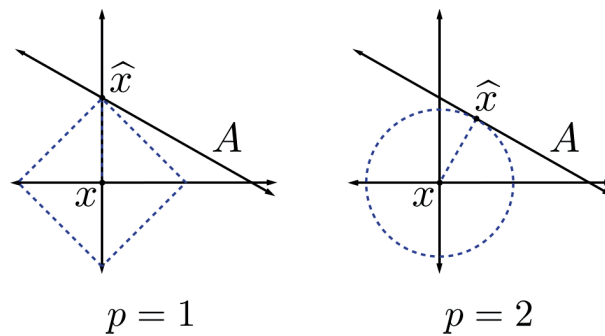


Figure 2.2: Illustration of the ℓ_p approximation \hat{x} of a point $x \in \mathbb{R}^2$ by a one-dimensional space A . Dashed lines represent an ℓ_1 ball and an ℓ_2 ball (set of vectors with the same ℓ_1 or ℓ_2 norm, respectively). *Illustration taken from [22]. Image used with permission.*

The theory of CS performs an analysis of problem (2.16) and provides recovery guarantees under certain conditions on the signal and the sensing matrix [23]. However, convex

optimization is particularly versatile and can account for variations in the formulation of the problem. For instance, the unconstrained version of the ℓ_1 minimization problem can also be considered, i.e.,

$$\min_{\hat{\alpha} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{y} - \Theta \hat{\alpha}\|_2^2 + \lambda \|\hat{\alpha}\|_1. \quad (2.17)$$

For some value of λ , this optimization problem will yield the same result as the formulation (2.16). Another possible variation implies optimizing the signal \mathbf{x} itself (analysis-based formulation), instead of solving the minimization problem for the representation vector $\hat{\alpha}$ and then recovering the signal through $\hat{\mathbf{x}} \equiv \Psi \hat{\alpha}$ (synthesis-based methods). Note that in the case of Ψ being an orthonormal basis the two approaches are equivalent.

2.4.2 Beyond ℓ_1 : iterative reweighted ℓ_1 minimization

In this subsection, we present a reweighted iterative algorithm that in many situations finds the good solution for (2.14) but with less measurement requirements than considering the ℓ_1 surrogate [24]. In this approach, the ℓ_0 norm in (2.15) is substituted by a weighted ℓ_1 norm defined as $\|\mathbf{w}\alpha\|_1 = \sum_i w_i |\alpha_i|$, for positive weights w_i . A reweighted scheme algorithm is defined, alternating between solving a problem of the form:

$$\hat{\alpha}^{(t)} = \min_{\hat{\alpha} \in \mathbb{C}^N} \|\mathbf{W}^{(t)} \hat{\alpha}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Theta \hat{\alpha}\|_2 \leq \epsilon, \quad (2.18)$$

and updating the weights w_i . In (2.18), matrix $\mathbf{W}^{(t)} \in \mathbb{R}^N$ represents a diagonal matrix with coefficients $\{w_1, \dots, w_N\}$ in the main diagonal and superindex t indexes the iteration number.

In this kind of scheme, large weights tend to *discourage* nonzero entries whereas small weights *promote* nonzero entries in the solution. Therefore, as it is extensively discussed in [24], the weights need to relate to the inverse of the associated element of the solution at the previous iteration, so as to lead to an ℓ_0 -norm prior at convergence. In [24], the authors propose to update the weights as follows:

$$w_i^{(t+1)} = \frac{1}{|x_i^{(t)}| + \tau} \quad (2.19)$$

and terminate either on convergence or when t reaches a maximum-allowed number of iterations. The parameter τ that appears in the definition of the weights has the role of providing stability and ensuring that all weights are well defined, even for zero coefficients.

With this definition (2.19), the weights compensate the fact that simple ℓ_1 minimization does not treat all coefficients evenly. Indeed, when minimizing the convex ℓ_1 norm to promote sparsity, larger coefficients are actually more penalized than the smaller ones, whereas they would be equally treated if one was directly minimizing the original ℓ_0 pseudo-norm.

The reweighted ℓ_1 minimization outperforms plain ℓ_1 minimization in a variety of setups and since the number of iterations is typically very low, its additional computational cost is affordable [24]. Illustration 2.3 provides an visual example of a situation where the (unweighted) ℓ_1 minimization fails to recover the correct solution whereas the weighted- ℓ_1 minimization, does not. In this simple 3D situation, the feasible set intersects the interior of the ℓ_1 ball centered in the origin with a radius equal to $\|x_0\|_1$, and therefore minimizing the ℓ_1 norm does not find the correct solution (situation (b) in Figure 2.3). However, *weighting* the ℓ_1 ball with appropriate weights provokes a “sharpening” effect on the shape of the ball, avoiding its intersection with the linear constraint and leading the weighted problem to find the correct solution.

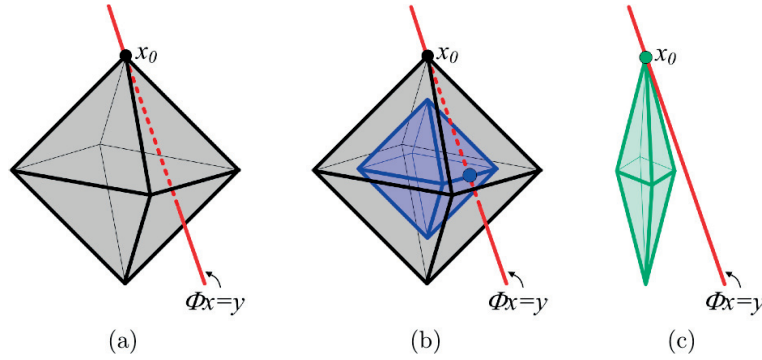


Figure 2.3: Visual example of sparse signal recovery using a weighted ℓ_1 norm. **(a):** The linear constraint $y = \Phi x$ and the ℓ_1 ball containing the sparse signal to be recovered x_0 are shown. **(b):** $x \neq x_0$ with $\|x\|_1 < \|x_0\|_1$ exists. **(c):** Weighted ℓ_1 ball. No x such that $\|Wx\|_1 < \|Wx_0\|_1$ exists. Illustration taken from [24]. Image used with permission.

2.4.3 Low-rank matrix recovery

The problem of recovering a matrix from some sample of its entries, or from fewer linear functionals about the matrix, is known as the matrix completion problem [25, 26, 27]. Mathematically, it can as well be formulated as a linear inverse problem

$$y = \mathcal{A}(X) + \eta, \quad (2.20)$$

where the unknown would be a matrix $X \in \mathbb{R}^{N_1 \times N_2}$ and \mathcal{A} a linear operator acting on the space of $N_1 \times N_2$ matrices. Note that matrix X is represented by $N_1 \cdot N_2$ numbers but it only has $(2N - r) \cdot r$ degrees of freedom, with $N = \max(N_1, N_2)$ and $r = \text{rank}(X)$. When the rank is small, this is notably smaller than $N_1 \cdot N_2$.

Analogously to the vector case, the assumption that the unknown has low dimension (low rank) makes the search of solutions feasible. Consequently, one would like to solve

the following optimization problem:

$$\min_{\mathbf{X} \in \mathbb{R}^{N_1 \times N_2}} \text{rank}(\mathbf{X}) \quad \text{subject to} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{X})\|_2 \leq \epsilon. \quad (2.21)$$

However, even if (2.21) appears as the most natural problem to solve, it is actually NP-hard.

A popular alternative to problem (2.21) resorts to minimizing the nuclear norm of the unknown, i.e. $f_r(\cdot) = \|\cdot\|_*$ [27], leading to the following formulation:

$$\min_{\mathbf{X} \in \mathbb{R}^{N_1 \times N_2}} \|\mathbf{X}\|_* \quad \text{subject to} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{X})\|_2 \leq \epsilon. \quad (2.22)$$

Interestingly, the nuclear norm is a convex functional and therefore (2.22) can be solved efficiently using convex optimization methods (see section 2.5). Note that, as already mentioned in section 2.2, whereas the nonconvex function $\text{rank}(\mathbf{X})$ corresponds to the ℓ_0 norm of the vector of singular values of \mathbf{X} , $\|\mathbf{X}\|_*$ can be seen as its ℓ_1 norm. This creates a strong analogy between the ℓ_1 prior in (2.16), appearing as the convex surrogate for the natural ℓ_0 minimization, and the nuclear norm of a matrix appearing as a convex alternative to its rank in (2.22).

This formulation is used in the so-called Phase Lift approach [28]. In that framework, quadratic measurements of the form $|\langle \mathbf{x}, \mathbf{a}_i \rangle|^2$ for given projection vectors \mathbf{a}_i , are seen as linear measurements on the rank-1 matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\dagger$ representing the outer product of the signal with itself.

2.5 Convex optimization and proximal splitting methods

A real valued function $f(\mathbf{x})$, from \mathbb{R}^N to \mathbb{R} , is called convex if

$$f((1 - \beta)\mathbf{x}_1 + \beta\mathbf{x}_2) \leq (1 - \beta)f(\mathbf{x}_1) + \beta f(\mathbf{x}_2) \quad (2.23)$$

for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$ and any $\beta \in [0, 1]$. Optimization problems including convex objective functions and convex constraints, called convex optimization problems, have many attractive properties, in particular the essential property that any local minimum must be a global minimum, which comes directly from the definition of a convex function. Also, convex problems can be efficiently solved, both in theory (i.e., via algorithms with worst-case polynomial complexity) and in practice [29].

Among the broad range of convex optimization methods, *proximal splitting methods*, exhaustively reviewed in [30], offer great flexibility and are shown to capture and extend several well-known algorithms in a unifying framework. Examples of proximal splitting algorithms include Douglas-Rachford, iterative thresholding, projected Landweber, projected gradient, forward-backward, alternating projections, alternating direction method

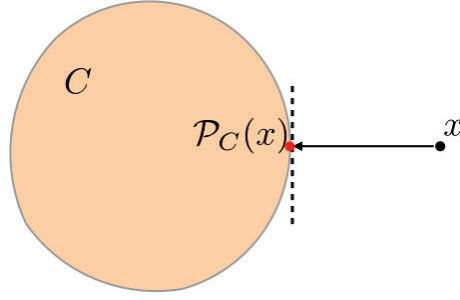


Figure 2.4: Projection onto a convex set.

of multipliers and alternating split Bregman [30]. They solve optimization problems of the form

$$\min_{\mathbf{x} \in \mathbb{R}^N} f_1(\mathbf{x}) + \dots + f_K(\mathbf{x}), \quad (2.24)$$

where $f_1(\mathbf{x}), \dots, f_K(\mathbf{x})$ are convex lower semicontinuous functions from \mathbb{R}^N to \mathbb{R} . In the case of convex constrained problems, they can be reformulated as unconstrained problems by using the indicator function of the convex constraint set as one of the functions in (2.24), i.e. $f_k(\mathbf{x}) = i_C(\mathbf{x})$ where C represents the convex constraint set. The indicator function, defined as $i_C(\mathbf{x}) = 0$ if $\mathbf{x} \in C$ or $i_C(\mathbf{x}) = +\infty$ otherwise, belongs to the class of convex lower semicontinuous functions. Note that complex-valued vectors are treated as real-valued vectors with twice the dimension accounting for real and imaginary parts [31].

Proximal splitting methods proceed by *splitting* the contribution of the functions $f_1(\mathbf{x}), \dots, f_K(\mathbf{x})$ individually so as to yield an easily implementable algorithm. They are called *proximal* because each function in (2.24) is involved by its proximity operator, which can be seen as a generalization of a convex projection operator.

Let f be a convex lower semicontinuous function from \mathbb{R}^N to \mathbb{R} , then the proximity operator of f is defined as:

$$\text{prox}_f(\mathbf{x}) \triangleq \arg \min_{\mathbf{z} \in \mathbb{R}^N} f(\mathbf{z}) + \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2. \quad (2.25)$$

In the case of indicator functions of convex sets, the proximity operator is the projection operator onto the set (see figure 2.4):

$$\mathcal{P}_C(\mathbf{x}) = \arg \min_{\mathbf{z} \in \mathbb{R}^N} i_C(\mathbf{z}) + \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2. \quad (2.26)$$

The proximity operator of the ℓ_1 norm is well-known as the soft-thresholding operator, or shrinkage:

$$\text{prox}_{\lambda \|\cdot\|_1}(\mathbf{x}) = \max\{|\mathbf{x}| - \lambda, \mathbf{0}\} \cdot \text{sign}(\mathbf{x}) \triangleq \text{soft}(\mathbf{x}, \lambda). \quad (2.27)$$

$\text{soft}(\mathbf{x}, \lambda)$ sends all components $x_i \leq \lambda$ to zero and shrinks the rest and consequently, it

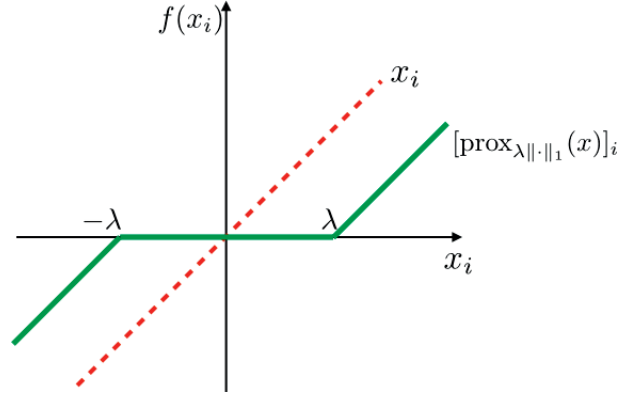


Figure 2.5: Soft-thresholding operator

induces sparsity to \mathbf{x} . Figure 2.5 shows the soft-thresholding function for one of the components x_i of a vector \mathbf{x} and can help the reader to intuitively see how the ℓ_1 regularization induces sparsity.

Most proximal splitting algorithms reach a solution to (2.24) by alternately applying the proximity operator associated with each function. For example, in the case that all functions in (2.24) are indicator functions, the algorithm reduces to the classical projection onto convex sets algorithm [29], which performs alternate projections to reach the solution. An important feature of proximal splitting methods is that they offer a powerful framework for solving convex problems in terms of speed and scalability of the techniques to very high dimensions. Hereafter two of the most popular proximal splitting methods, the forward-backward algorithm and the Douglas-Rachford algorithm, are described in detail since they are referred to in parts I and II of the manuscript. See [30] for a complete review of proximal splitting methods and their applications in signal and image processing.

2.5.1 Forward-Backward algorithm

This algorithm solves optimization problems of the form:

$$\min_{\mathbf{x} \in \mathbb{R}^N} f_1(\mathbf{x}) + f_2(\mathbf{x}), \quad (2.28)$$

with f_1 being a lower semicontinuous convex function from \mathbb{R}^N to \mathbb{R} and such that $\text{dom} f \neq \emptyset$, $f_2 : \mathbb{R}^N \rightarrow \mathbb{R}$ being convex with a β -Lipschitz continuous gradient ∇f_2 , i.e.

$$\|\nabla f_2(\mathbf{x}) - \nabla f_2(\mathbf{y})\| \leq \beta \|\mathbf{x} - \mathbf{y}\|, \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N, \beta > 0, \quad (2.29)$$

and assuming that $f_1(\mathbf{x}) + f_2(\mathbf{x}) \rightarrow +\infty$ as $\|\mathbf{x}\| \rightarrow +\infty$.

Starting from an initial point $\mathbf{x}_0 \in \mathbb{R}^N$, for a chosen step-size parameter $\gamma \in (0, \min\{1, 1/\beta\})$ and iterating as follows:

$$\mathbf{x}_{n+1} = \underbrace{\text{prox}_{\gamma_n f_1}}_{\text{backward step}} \left(\underbrace{\mathbf{x}_n - \gamma_n \nabla f_2(\mathbf{x}_n)}_{\text{forward step}} \right), \quad (2.30)$$

a sequence $(x_n)_{n \geq 0}$ is generated that converges to a solution to problem (2.28) [30]. Equation (2.30) summarizes the forward-backward algorithm, that requires one single proximal step at each iteration.

2.5.2 Douglas-Rachford algorithm

This algorithm provides solutions for problems of the form (2.28) where the Lipschitz-differentiability condition on f_2 is relaxed. Starting from $\mathbf{y}_0 \in \mathbb{R}^N$ and choosing parameters $\gamma > 0, \epsilon \in (0, 1)$, the Douglas-Rachford algorithm can be summarized in the following 2-steps proximal scheme:

1. $\mathbf{x}_n = \text{prox}_{\gamma f_2} \mathbf{y}_n$
2. $\mathbf{y}_{n+1} = \mathbf{y}_n + \text{prox}_{\gamma f_1}(2\mathbf{x}_n - \mathbf{y}_n) - \mathbf{x}_n$

The Douglas-Rachford algorithm can be seen as a generalization of the forward-backward algorithm since it does not require the functions involved to be Lipschitz-differentiable. However, it is numerically more complex than the latter since it involves two proximal steps at each iteration.

2.6 Conclusion

In this chapter, we have presented an overview on linear inverse problems since the theory of regularization methods is very well studied for them and they can be easily solved using the versatile framework of convex optimization. We have focused on sparse inverse problems, since most of the natural signals are known to be sparse or compressible in a suitable domain. The theory of Compressed Sensing offers a powerful framework for sparse signal recovery. We have reviewed its main results and formulated different convex minimization problems that leverage some kind of sparsity or low dimensionality of the signal. At the reconstruction level, we have given some notions on convex optimization and have revised two popular proximal splitting methods that we will use further in this thesis.

Before concluding, we want to highlight that convex optimization is an extremely adaptable framework that enables to easily include prior information about the signal, such as positivity, as long as it is formulated as a convex constraint. Therefore, different minimization problems, other than the ones described in this chapter, can be straightforwardly defined for the recovery. For example, a TV norm* may also be substituted for the

*The TV norm of a signal is simply defined as the ℓ_1 norm of the magnitude of its gradient

ℓ_1 norm in (2.16) for signals with sparse or compressible gradients. All this flexibility in the definition of the optimization problem is an important manifestation of the versatility of the convex optimization scheme and, as the reader will see, we take advantage of it when designing the novel algorithms that we present in subsequent chapters. In particular, in chapter 3 we propose a generalization of the Phase Lift approach (2.4.3) for tensor recovery. In chapters 5 and 6, we have designed novel reweighting schemes to minimize a weighted ℓ_1 norm (2.4.2) that induces structure on the sparsity of the solution thanks to a specific definition of the weights.

Part of this chapter was published in [3].

Part I

Optical Interferometry

Tensor optimization for optical-interferometric imaging

3.1 Introduction

In interferometry, electromagnetic waves are superposed to retrieve information from their emitting sources that otherwise would not be easily accessible. An astronomical interferometer consists of an array of telescopes that brings together their signals, offering a much higher resolution (approximately equivalent to that of a telescope of diameter equal to the largest separation between its individual elements).

The measurements associated with a given pair of telescopes (j_1, j_2) at one instant of observation t are known as the complex visibilities, $V_{j_1, j_2}(t)$. These measurements correspond to the Fourier transform of the image of interest at a spatial frequency $\nu_{j_1, j_2, m}(t) = \frac{\vec{B}}{\lambda} \equiv (u, v)$, where vector \vec{B} indicates the projection onto the plane of the sky of the baseline between the pair of telescopes (j_1, j_2) and λ denotes the wavelength. The two-dimensional space of spatial frequencies is commonly known as the (u, v) -plane.

What an interferometer actually measures are the complex visibilities averaged during a finite exposure duration:

$$\langle V_{j_1, j_2}(t) \rangle_m \approx G_{j_1, j_2, m} \hat{I}(\nu_{j_1, j_2, m}). \quad (3.1)$$

In (3.1), $\nu_{j_1, j_2, m}$ denotes to the spatial frequency sampled by the pair of telescopes (j_1, j_2) , averaged over the m th exposure. $G_{j_1, j_2, m}$ indicates the effective optical transfer function (OTF) and $\hat{I}(\nu)$ corresponds to the Fourier transform of $I(\theta)$, the brightness distribution of the observed object under a view angle θ . The reader can find a visual representation of an optical interferometer and its main associated parameters in figure 3.1.

At radio wavelengths, the OTF is nonnegligible ($G_{j_1, j_2, m} \neq 0$) and the visibilities in (3.1) are indeed accessible, thereby setting a sparse Fourier inverse problem in the perspective of image recovery. When the OTF can be calibrated, the problem boils down to a deconvolution problem with sparse Fourier data. Several methods have been proposed to solve such an ill-posed problem, for instance, the standard CLEAN algorithm operates by local iterative removal of the convolution kernel associated with the partial Fourier

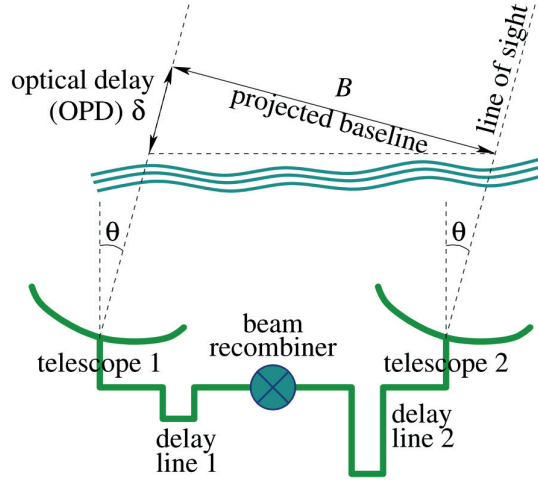


Figure 3.1: Sketch of an optical interferometer. The projected baseline is denoted by B ; θ indicates the view angle. *Illustration taken from [32]. Used with permission*

coverage [33]. Alternatively, convex optimization methods regularizing the inverse problem through sparsity constraints have recently been proposed in the framework of the recent theory of compressive sampling [34, 35, 36, 37, 38, 39, 40].

At optical wavelengths though, atmospheric turbulence induces a random phase delay that drives $G_{j_1, j_2, m} \approx 0$, implying a systematic cancellation of the visibility values. To overcome this challenge current optical interferometers can retrieve power spectrum information:

$$S_{j_1, j_2, m} = \langle |V_{j_1, j_2}(t)|^2 \rangle_m \approx H_{j_1, j_2, m} |\hat{I}(\nu_{j_1, j_2, m})|^2. \quad (3.2)$$

In this case, $H_{j_1, j_2, m}$ becomes a nonzero transfer function that can be easily estimated. Since the power spectrum measurements do not provide any phase information, the bispectrum of the complex visibilities for a triplet of telescopes (j_1, j_2, j_3) can also be measured [32, 41, 42]:

$$B_{j_1, j_2, j_3, m} = \langle V_{j_1, j_2}(t) V_{j_2, j_3}(t) V_{j_3, j_1}(t) \rangle_m \approx J_{j_1, j_2, j_3, m} \hat{I}(\nu_{j_1, j_2, m}) \hat{I}(\nu_{j_2, j_3, m}) \hat{I}(\nu_{j_3, j_1, m}). \quad (3.3)$$

The transfer function $J_{j_1, j_2, j_3, m}$ in (3.3) can also be easily estimated and, since it takes real values, has no effect on the phase of the bispectrum, the so-called *phase closure*. Note that the phase closure actually corresponds to the sum of three phases around a closed triangle of baselines and indeed is a very robust measure insensitive to atmosphere-induced phase shifts [43]. However, for any number of telescopes, there are always less independent Fourier phases, among all the possible phase closures that can be measured, than the actual number of phases we would like to determine [43]. These considerations apply both to aperture masking interferometry on a single telescope [44, 45, 46], as well

to optical interferometer arrays such as the Very Large Telescope Interferometer (VLTI)*. As a result, the problem of image recovery in optical interferometry represents a very challenging ill-posed nonlinear Fourier inverse problem with incomplete phase information. The simplified version of this problem, with calibrated and debiased power spectrum and bispectrum measurements, reads as:

$$\begin{aligned} S_{j_1,j_2,m} &= |\hat{I}(\nu_{j_1,j_2,m})|^2 + S_{j_1,j_2,m}^{err}, \\ B_{j_1,j_2,j_3,m} &= \hat{I}(\nu_{j_1,j_2,m})\hat{I}(\nu_{j_2,j_3,m})\hat{I}(\nu_{j_3,j_1,m}) + B_{j_1,j_2,j_3,m}^{err}, \end{aligned} \quad (3.4)$$

where $S_{j_1,j_2,m}^{err}$ and $B_{j_1,j_2,j_3,m}^{err}$ correspond to noise terms.

To solve (3.4), the state-of-the-art MiRA method [47] takes a maximum a posteriori (MAP) approach where the image is the solution of an optimization problem with an objective function $f(\mathbf{x}) = f_{\text{data}}(\mathbf{x}) + \ell f_{\text{prior}}(\mathbf{x})$, for some arbitrary parameter ℓ to be tuned, and with additional positivity and total flux constraints. Sparsity priors have in particular been promoted [47, 48]. The data nonlinearity induces nonconvexity of the objective function. The adopted strategy is to perform only local optimization, in the context of which the solution depends not only on the data and on the priors but also strongly on the initial image and on the path followed by the local optimization method. The WISARD alternative [49] takes a two-step alternate minimization self-calibration approach. Firstly, the missing Fourier phases are recovered on the basis of a current estimate and phase closure information enabling to build pseudocomplex visibilities. Secondly, the image is recovered from the pseudocomplex visibilities as in radio interferometry. While the second step is convex and leads to a unique image independently of the initialization, the first step is not. The overall procedure remains nonconvex and the final solution depends on the initial guess. In summary, state-of-the-art methods are nonconvex due to the intrinsic data nonlinearity [32], and therefore known to suffer from a strong sensitivity to initialization.

The approaches proposed in this work stem from a different perspective. We firstly formulate a linear version of the problem for the real and positive supersymmetric rank-1 order-3 tensor $\mathcal{X} = \mathbf{x} \circ \mathbf{x} \circ \mathbf{x}$ formed by the tensor product of the size- N vector \mathbf{x} representing the image under scrutiny[†] with itself. This allows us to pose a linear convex problem for recovery of a size- N^3 tensor \mathcal{X} with built-in supersymmetry. We regularize the inverse problem through a nuclear norm relaxation of a low-rank constraint, also enforcing reality, positivity and optionally sparsity constraints. We also study a different nonlinear nonconvex approach with built-in rank-1 constraint but where supersymmetry is relaxed, formulating the problem for the tensor product $\mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3$ of 3 size- N vectors. In contrast with the state of the art though, only linear convex minimization subproblems are solved,

*www.eso.org/sci/facilities/paranal/telescopes/vlti/

[†]The image \mathbf{x} correspond to the vector of coefficients of the brightness distribution in a representation basis. In this work we always represent the image in real space, therefore \mathbf{x} corresponds directly to the unfolding of I

alternately and iteratively for the vectors, also enforcing reality and positivity[‡]. While the former approach is much heavier than the latter in terms of memory requirements and computation complexity due to the drastically increased dimensionality of the unknown, the underlying convexity ensures essential properties of convergence to a global minimum of the objective function and independence to initialization, justifying a comparative analysis. For numerical experiments, we consider a generic discrete measurement setting where measurements identify with triple products of discrete Fourier coefficients of \mathbf{x} . These triple products are selected randomly according to a variable-density scheme sampling more densely low spatial frequencies, and are affected by simple additive Gaussian noise.

The rest of the chapter is organized as follows: In subsection 3.2.1, we introduce our generic discrete data model and describe our new linear tensor formulation of the optical-interferometric imaging problem. In subsections 3.2.2 and 3.2.3, the new AM, NM and NM-RW approaches are discussed. Our simulation setting for comparison of these two methods and corresponding results are presented in section 3.3. Section 3.4 concludes this chapter with a reminder of our contributions as well as a mention to future work.

3.2 Materials and Methods

3.2.1 Data model and tensor formulation

For the sake of simplicity, we adopt a discrete setting where the intensity image of interest is represented by the real and positive vector $\mathbf{x} \in \mathbb{R}_+^N$ with components x_i . Its 2D discrete Fourier transform is denoted $\hat{\mathbf{x}} \in \mathbb{C}^N$ with components \hat{x}_i . By abuse of notation, we denote \hat{x}_{-i} the component of $\hat{\mathbf{x}}$ at the opposite spatial frequency to that associated with \hat{x}_i . Signal reality implies $\hat{x}_{-i} = \hat{x}_i^*$, where $*$ stands for complex conjugation.

The optical interferometry inverse problem is simplified considering a generic discrete measurement setting where the closure constraint is relaxed and optical-interferometric measurements take the generic form of a triple product of Fourier coefficients of the image: $\hat{x}_i \hat{x}_j \hat{x}_k$. Power spectrum measurements follow with $j = -i$, and $k = 0$ (\hat{x}_0 stands for the Fourier coefficient at zero frequency), and explicit bispectrum measurements would follow from the constraint that the spatial frequencies associated with \hat{x}_i , \hat{x}_j , and \hat{x}_k sum to zero. In this context, measurements are performed on the frequencies of a discrete grid in the Fourier plane, the so-called *frequels*. In a real scenario the Fourier transform should be evaluated at (non-equispaced) continuous frequencies [31]. We write the measurement equation in compact form as

$$\mathbf{y} = \mathcal{V}(\mathbf{x}) + \mathbf{n}, \quad (3.5)$$

where \mathcal{V} is a nonlinear operator providing an undersampled set of triple products of Fourier

[‡]We also attempted an alternative nonconvex approach consisting in solving the nonlinear problem directly for \mathbf{x} , using the nonconvex projected gradient method proposed by [50]. First simulations did not produce any meaningful reconstruction and this approach was discarded.

coefficients of \mathbf{x} . The measurement vector $\mathbf{y} \in \mathbb{C}^M$, with components y_a ($1 \leq a \leq M$) is assumed to be affected by a simple noise vector $\mathbf{n} \in \mathbb{C}^M$ with i.i.d. Gaussian components n_a . The number of measurements is typically smaller than the signal dimension: $M < N$. Finally, we assume that the total flux is measured independently and consider a normalized signal such that $\sum_i x_i = \hat{x}_0 = 1$. This flux normalization is approximately enforced by adding the data point $\hat{x}_0^3 = 1$.

In what follows, we show how to bring the linearity of the measurement scheme by lifting the image model from a vector to a tensor formulation. The reader can review some tensor definitions and notations in section 2.2. The measurement model (3.5) can be recast as the following linear model for the real and positive supersymmetric rank-1 order-3 tensor $\mathcal{X} = \mathbf{x} \circ \mathbf{x} \circ \mathbf{x} \in \mathbb{R}_+^{N \times N \times N}$:

$$\mathbf{y} = \mathcal{T}(\mathcal{X}) + \mathbf{n}, \quad (3.6)$$

where the linear operator \mathcal{T} consists in performing a 2D discrete Fourier transform along each of the 3 dimensions, identified by an operator \mathcal{F} , followed by a selection and vectorization operator \mathcal{M} providing variable-density undersampling in this 6D Fourier space: $\mathcal{T} = \mathcal{M}\mathcal{F}$. The unit flux measurement is also included in the mask as a measurement on the “triple-zero frequency”. Note that this formulation is a generalization of the Phase Lift approach for the well-known phase retrieval problem [28]. In that framework, quadratic measurements of the form $|\langle \mathbf{x}, \mathbf{a}_i \rangle|^2$ for given projection vectors \mathbf{a}_i , are seen as linear measurements on the rank-1 matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\dagger$ representing the outer product of the signal with itself.

We note however that the rank-1 and supersymmetry properties are not explicitly built-in in the tensor formulation (3.6), which thereby presents a drastically increased dimensionality, N^3 , of the unknown \mathcal{X} compared to the original \mathbf{x} of size N in (3.5). In the following sections, we discuss our two different regularization schemes for tensor recovery. We firstly study a nonconvex alternate minimization (AM) approach where the rank-1 constraint is built-in, and subsequently move to a convex nuclear minimization (NM) scheme with built-in supersymmetry.

3.2.2 Rank-1 alternate minimization (AM)

3.2.2.1 Algorithm formulation

We consider the following explicit rank-1 formulation of (3.6), where supersymmetry is relaxed:

$$\mathbf{y} = \mathcal{T}(\mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3) + \mathbf{n}. \quad (3.7)$$

The measurements can now be understood as an undersampled set of products of Fourier coefficients of \mathbf{u}_1 , \mathbf{u}_2 , and \mathbf{u}_3 , thus bringing back nonlinearity. We consider the following

nonconvex minimization problem for tensor recovery:

$$\min_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \in \mathbb{R}_+^N} \|\mathcal{T}(\mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3) - \mathbf{y}\|_2^2. \quad (3.8)$$

A priori this problem seems as nonlinear and nonconvex as the initial problem (3.5). Thanks to the nonsupersymmetric relaxation though, an alternate minimization algorithm can be designed, solving sequentially for each variable (\mathbf{u}_1 , \mathbf{u}_2 or \mathbf{u}_3) while keeping the other two fixed, and iterating until convergence. At each iteration, the 3 linear and convex subproblems

$$\min_{\mathbf{u}_p \in \mathbb{R}_+^N} \|\mathcal{T}_{(\mathbf{u}_q \mathbf{u}_s)} \mathbf{u}_p - \mathbf{y}\|_2^2, \quad (3.9)$$

are therefore solved sequentially for $1 \leq p \neq q \neq s \leq 3$, where the linear operators $\mathcal{T}_{(\mathbf{u}_q \mathbf{u}_s)}$ are defined by $\mathcal{T}_{(\mathbf{u}_q \mathbf{u}_s)} \mathbf{u}_p \equiv \mathcal{T}(\mathbf{u}_p \circ \mathbf{u}_q \circ \mathbf{u}_s)$. In each subproblem the linear operator is computed using the values of the fixed variables at the current step. The final AM algorithm is depicted in Algorithm 3. The algorithm is initialized with the same random vector for each of the 3 subproblems. The algorithm is stopped when the relative variation between the objective function in (3.8) evaluated at successive solutions is smaller than some predefined bound or after the maximum number of iterations allowed is reached. At convergence, the tensor solution takes the form of 3 vectors \mathbf{u}_1 , \mathbf{u}_2 , and \mathbf{u}_3 . We have no guarantee that the 3 solution vectors are identical and heuristically choose the final solution to be their mean as shown in step 8 of Algorithm 3[§].

Algorithm 1 AM algorithm

- 1: Initialize $k = 1$, $\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}, \mathbf{u}_3^{(0)} \in \mathbb{R}^N$.
 - 2: **while** not converged **do**
 - 3: $\mathbf{u}_1^{(k)} = \arg \min_{\mathbf{u}_1} \|\mathcal{T}_{(\mathbf{u}_2^{(k-1)} \mathbf{u}_3^{(k-1)})} \mathbf{u}_1 - \mathbf{y}\|_2^2$.
 - 4: $\mathbf{u}_2^{(k)} = \arg \min_{\mathbf{u}_2} \|\mathcal{T}_{(\mathbf{u}_1^{(k)} \mathbf{u}_3^{(k-1)})} \mathbf{u}_2 - \mathbf{y}\|_2^2$.
 - 5: $\mathbf{u}_3^{(k)} = \arg \min_{\mathbf{u}_3} \|\mathcal{T}_{(\mathbf{u}_1^{(k)} \mathbf{u}_2^{(k)})} \mathbf{u}_3 - \mathbf{y}\|_2^2$.
 - 6: $k \leftarrow k + 1$
 - 7: **end while**
 - 8: $\mathbf{x}_{\text{AM}} = \frac{1}{3}(\mathbf{u}_1^{(k)} + \mathbf{u}_2^{(k)} + \mathbf{u}_3^{(k)})$
 - 9: **return** \mathbf{x}_{AM}
-

[§]Note that [51] prove that this alternate minimization approach converges to a critical point of the objective function (3.8), provided that terms of the form $\gamma \|\mathbf{u}_p - \bar{\mathbf{u}}_p\|_2^2$ controlling the distance between the current unknown \mathbf{u}_p with respect to its value at the previous iteration $\bar{\mathbf{u}}_p$ are added to the objective function in (3.9), for any $\gamma > 0$. Simulations in the context of the setting described in Section 3.3 show that the algorithm converges to the same solution for $\gamma \neq 0$ and $\gamma = 0$. Other simulations also show that starting the minimization of the three variables with the same random initial point leads to very similar solutions for the 3 vectors, or for their mean, both in terms of signal-to-noise ratio and visual quality.

3.2.2.2 Optimization details

To solve each of the subproblems in Algorithm 3 (steps 3–5) we resort to a forward-backward (projected gradient) algorithm [30]. The forward-backward algorithm solves (3.9) using a two step procedure: a gradient descent step (forward step) to minimize the quadratic function in (3.9), and a projection step (backward step) to bring back the current update to the constraint set. The algorithm uses the following recursion:

$$\mathbf{u}_p^{(t+1)} = \text{prox}_{i_C} \left(\mathbf{u}_p^{(t)} + \mu_p^{(t)} \mathbb{T}_{(u_q u_s)}^\dagger \left(\mathbf{y} - \mathbb{T}_{(u_q u_s)} \mathbf{u}_p^{(t)} \right) \right), \quad (3.10)$$

where t denotes the iteration variable, $C = \mathbb{R}_+^N$ and $\mu^{(t)}$ is a variable step size that controls the gradient descent update. The step size is adapted using a backtracking line-search procedure [52]. The proximity operator prox_{i_C} is nothing but the projector onto the positive orthant \mathbb{R}_+^N , i.e. setting the imaginary part and the negative values of the real part to zero [29].

The memory requirement to solve this minimization problem is dominated by the storage of the 3 vectors, which is of size $\mathcal{O}(N)$. In terms of computation time, the algorithm is dominated at it each iteration by the application of the operator \mathbb{T} which computes 3 2D FFTs of size N , with an asymptotic complexity of order $\mathcal{O}(N \log N)$. This approach is computationally efficient. In contrast with the state-of-the-art approaches such as MiRA and WISARD, it brings convexity to the subproblems. But the global problem remains nonconvex and the solution may still depend on the initialization. One can easily identify convergence to a local minimum through large residual values of the objective function. With the aim to mitigate the dependence to initialization, and as suggested by [53], we propose to run the algorithm n_{ri} times with random initializations, choosing *a posteriori* the solution with minimum objective function value.

3.2.3 Supersymmetric nuclear minimization (NM)

3.2.3.1 Algorithm formulation

Tensor supersymmetry can be embedded in various ways. One approach is to formulate the inverse problem (3.6) only for the subset of variables \mathcal{X}_{ijk} with $i \leq j \leq k$. The collection of these values define the “subtensor” \mathcal{X}_s , which can be related to \mathcal{X} by an operator \mathcal{R} replicating tensor components over all permutations for each triplet (i, j, k) : $\mathcal{X} = \mathcal{R}(\mathcal{X}_s)$. The inverse problem would thus read $\mathbf{y} = [\mathbb{T}\mathcal{R}](\mathcal{X}_s) + \mathbf{n}$. We adopt an alternative and equivalent approach consisting in substituting the original measurement vector \mathbf{y} by its replicated version $\mathcal{R}(\mathbf{y})$, and using a symmetrized version \mathcal{M}_s of the selection mask, ensuring that all permutations of a triplet (i, j, k) are assumed to be measured. We will see below why a symmetrized data vector together with a symmetrized measurement operator represent a sufficient condition to impose the tensor symmetry at each step of the algorithm in our approach, and in particular supersymmetry of the solution. The modified

inverse problem thus reads as:

$$\mathbf{y}_s = \mathcal{T}_s(\mathcal{X}) + \mathbf{n}_s, \quad (3.11)$$

with $\mathbf{y}_s = \mathcal{R}(\mathbf{y})$, $\mathbf{n}_s = \mathcal{R}(\mathbf{n})$ and $\mathcal{T}_s = \mathcal{M}_s \mathcal{F}$ denoting the symmetrized versions of the measurement vector, noise vector and measurement operator, respectively. Without loss of generality, we assume that the initial selection operator \mathcal{M} contains no redundant measurements, i.e. $i \leq j \leq k$. This ensures that \mathcal{R} is well-defined. Also note that the noise statistics remains unaltered and only concerns the entries before replication.

Low-rankness, reality and positivity will be imposed as regularization priors in the convex minimization problem to be defined. As pointed out, the rank of a tensor is difficult to handle since the problem of finding $\text{rank}(\mathcal{X})$ is NP-hard. Computing the rank of different matricization of the tensor is an easier task. The unfoldings of a rank-1 tensor are actually rank-1 matrices, so that a low n -rank constraint can be used as a proxy for low-rankness. The rank of a matrix is however a nonconvex function. The nuclear norm, defined as the ℓ_1 -norm of its singular values, is a well-known convex relaxation of the rank function that was recently promoted in matrix recovery theory [25]. Building on those results, [54] tackle the low- n -rank tensor recovery problem through the minimization of the sum of the nuclear norms of the mode- n matricizations $\mathbf{X}_{(n)}$ for all n . In the supersymmetric case, the mode- n matricizations are all identical and denoted $\mathbf{X}_{(n)} = \mathcal{U}(\mathcal{X}) \in \mathbb{C}^{N \times N^2}$, where \mathcal{U} stands for the unfolding operator. We propose here to exploit the symmetry of the tensor under scrutiny, together with the signal normalization, to promote a computationally more efficient low-rank prior. Relying on these properties, we note that summations over one index of a tensor of the form $\mathbf{x} \circ \mathbf{x} \circ \mathbf{x}$ with $\sum_i x_i = 1$ leads to the order-2 tensor $\mathbf{x} \circ \mathbf{x}$, which is real, positive, symmetric, as well as rank-1 and positive-semidefinite. We define \mathcal{C} as the operator performing the summation over one dimension. Once more supersymmetry ensures that the resulting matrix is independent of the choice of the dimension along which components are summed up: $\mathcal{C}(\mathcal{X}) \in \mathbb{C}^{N \times N}$ with $[\mathcal{C}(\mathcal{X})]_{ij} = \sum_k \mathcal{X}_{ijk}$. A low-rank constraint on $\mathcal{C}(\mathcal{X})$ will be promoted, through a nuclear norm minimization, as a convex proxy for the low-rankness of \mathcal{X} . Positive-semidefiniteness of $\mathcal{C}(\mathcal{X})$, i.e. positivity of the eigenvalues, which are then identical to the singular values, may also be explicitly added as a convex prior, denoted $\mathcal{C}(\mathcal{X}) \succeq 0$, together with the convex reality and positivity constraints of \mathcal{X} : $\mathcal{X} \in \mathbb{R}_+^{N \times N \times N}$. This summation approach is *a priori* computationally significantly more efficient given the reduced matrix size of $\mathcal{C}(\mathcal{X})$ compared to that of the unfolded matrix $\mathcal{U}(\mathcal{X})$.

The resulting convex nuclear norm minimization problem (NM) for \mathcal{X} thus reads as:

$$\min_{\mathcal{X} \in S} \|\mathcal{C}(\mathcal{X})\|_* \quad \text{subject to} \quad \|\mathbf{y}_s - \mathcal{T}_s(\mathcal{X})\|_2 \leq \epsilon, \quad (3.12)$$

where $S = S_1 \cap S_2$, with $S_1 = \mathbb{R}_+^{N \times N \times N}$ and $S_2 = \{\mathcal{X} \mid \mathcal{C}(\mathcal{X}) \succeq 0\}$. Recalling that the measurements \mathbf{y} are assumed to be corrupted with simple i.i.d. complex Gaussian noise with

variance $\sigma_n^2/2$ on real and imaginary parts, the residual estimator $\|\mathbf{y} - \mathcal{T}(\mathcal{X})\|_2^2$ follows a χ^2 distribution with $2M$ degrees of freedom, with expectation $2M$ and standard deviation is $(4M)^{1/2}$. For a large number of degrees of freedom the distribution is extremely peaked around its expectation value. This fact is related to the well-known phenomenon of the concentration of measure [38]. The value $\epsilon_0^2 = (2M + 4\sqrt{M})\sigma_n^2/2$, i.e. 2 standard deviations above the expectation, represents a high percentile of the distribution (in practice extremely close to $2M$), and consequently a likely bound for $\|\mathbf{n}\|_2^2$. An equivalent bound for the symmetrized residual noise term $\|\mathbf{y}_s - \mathcal{T}_s(\mathcal{X})\|_2^2$ may simply be inferred as $\epsilon^2 \simeq \alpha\epsilon_0^2$, where α is simply the ratio of number of components in \mathbf{y}_s and \mathbf{y} . We take the value $\alpha = 6$ as the relative number of (i, j, k) triplets with repeated indices in the mask is very small. Note that this last consideration only arises from the discrete setting adopted.

Once the tensor solution \mathcal{X}_{NM} is recovered, the problem of extracting the sought signal \mathbf{x}_{NM} remains. If the tensor solution was actually a real positive rank-1 supersymmetric tensor whose elements sum up to unity, the retrieval of \mathbf{x}_{NM} could be done in different ways, such as directly extracting the first eigenvector of matrix $\mathcal{C}(\mathcal{X}_{\text{NM}})$ or simply performing a sum over two dimensions $\sum_{jk} [\mathcal{X}_{\text{NM}}]_{ijk}$. The nuclear norm minimization approach however does not guarantee that the final solution is indeed rank-1. We therefore resort to the generic algorithm proposed in [55] to find the best rank-1 supersymmetric approximation $\mathcal{P}_1(\mathcal{X}_{\text{NM}})$ of a supersymmetric tensor \mathcal{X}_{NM} in the least square sense. This algorithm is a generalization for the tensor case of the power method applied to find the dominant eigenvector of matrices [56]. It boils down to determining a unitary vector \mathbf{x} and a scalar λ , such that $\|\mathcal{X} - \lambda \mathbf{x} \circ \mathbf{x} \circ \mathbf{x}\|$ is minimized, where $\|\cdot\|$ indicates simply the sum of the square of the components of the tensor. We denote the resulting solution as

$$\mathbf{x}_{\text{NM}} = [\mathcal{E}\mathcal{P}_1](\mathcal{X}_{\text{NM}}), \quad (3.13)$$

where \mathcal{E} formally represents the operator retrieving from a supersymmetric rank-1 order-3 tensor its underpinning vector. Note that this vector extraction problem is not convex [¶].

The final NM algorithm is shown in Algorithm 2. To solve the complex optimization problem in (3.12) we use the Douglas-Rachford splitting algorithm, which is tailored to solve problems of the form in (2.24) with $K = 2$. The problem in (3.12) can be reformulated as in (2.24) by setting $f_1(\mathcal{X}) = \|\mathcal{C}(\mathcal{X})\|_* + i_S(\mathcal{X})$ and $f_2(\mathbf{x}) = i_{C_\epsilon}(\mathbf{x})$, where $C_\epsilon = \{\mathcal{X} \in \mathbb{C}^{N \times N \times N} : \|\mathbf{y}_s - \mathcal{T}_s(\mathcal{X})\|_2 \leq \epsilon\}$. The main recursion of the Douglas-Rachford algorithm is detailed in steps 3-4 of Algorithm 2, where $\nu > 0$ and $\tau_k \in (0, 2)$ are convergence parameters. The sequence $\{\mathcal{X}^{(k)}\}$ generated by the recursion in Algorithm 2 converges to a solution of the problem (3.12) [30]. The algorithm is stopped when the relative variation between successive solutions, $\|\mathcal{X}^{(k)} - \mathcal{X}^{(k-1)}\| / \|\mathcal{X}^{(k-1)}\|$, is smaller than some

[¶]Note that in [55] a proof of convergence of their algorithm only for even-order tensors is provident. Simulations in the context of the setting described in Section 3.3 show that this procedure systematically converges for our order-3 tensors, and provides significantly better results than a heuristic procedure based on extracting the first eigenvector of $\mathcal{C}(\mathcal{X}_{\text{NM}})$ or performing a sum over two dimensions $\sum_{jk} [\mathcal{X}_{\text{NM}}]_{ijk}$.

bound $\xi \in (0, 1)$, or after the maximum number of iterations allowed, T_{\max} , is reached. In our implementation we use the values $\tau_k = 1, \forall t, \xi = 10^{-3}$ and $\nu = 10^{-1}$. In the following subsection we detail the computation of the proximity operators for f_1 and f_2 .

Algorithm 2 NM algorithm

```

1: Initialize  $k = 1, \mathcal{X}^{(1)} \in \mathbb{R}^{N \times N \times N}, \tau_k \in (0, 2)$  and  $\nu > 0$ .
2: while not converged do
3:    $\mathcal{Z}^{(k)} = \text{prox}_{\nu f_2}(\mathcal{X}^{(k)})$ .
4:    $\mathcal{X}^{(k+1)} = \mathcal{X}^{(k)} + \tau_k (\text{prox}_{\nu f_1}(2\mathcal{Z}^{(k)} - \mathcal{X}^{(k)}) - \mathcal{Z}^{(k)})$ .
5:    $k \leftarrow k + 1$ 
6: end while
7:  $\mathbf{x}_{\text{NM}} = [\mathcal{EP}_1](\mathcal{X}^{(k)})$ .
8: return  $\mathbf{x}_{\text{NM}}$ 

```

3.2.3.2 Optimization details

The computation of the proximal operator of f_1 , which includes the nuclear norm prior, as well as the positive-semidefiniteness, reality and positivity constraints, is itself a complicated optimization problem. Therefore the dual forward-backward algorithm [30] is used at each iteration of the Douglas-Rachford recursion to compute the proximal operator of f_1 . We can decompose f_1 as $f_1(\mathcal{X}) = g_1(\mathcal{X}) + g_2(\mathcal{X})$, where $g_1(\mathcal{X}) = \|\mathcal{C}(\mathcal{X})\|_* + i_{S_1}(\mathcal{X})$ and $g_2(\mathcal{X}) = i_{S_2}(\mathcal{X})$. Let $\mathbf{Q}^{(0)} \in \mathbb{C}^{N \times N}$ and $\mathcal{S}^{(0)} \in \mathbb{C}^{N \times N \times N}$ be the all zero matrix and the all zero tensor respectively. The dual forward-backward algorithm uses the following recursion to compute $\text{prox}_{\nu f_1}(\mathcal{X})$:

$$\begin{aligned} \mathbf{Q}^{(t+1)} &= (\mathbf{I} - \text{prox}_{\nu g_1}) \left(\mathbf{Q}^{(t)} + \gamma_t \mathcal{C}(\mathcal{S}^{(t)}) \right) \\ \mathcal{S}^{(t+1)} &= \text{prox}_{\nu g_2} \left(\mathcal{X} - \mathcal{C}^\dagger(\mathbf{Q}^{(t+1)}) \right), \end{aligned} \quad (3.14)$$

where $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity operator and $\gamma_t \in (0, 2)$ is a step size. The sequence $\{\mathcal{S}^{(t)}\}$ converges linearly to $\text{prox}_{\nu f_1}(\mathcal{X})$.

The computation of $\text{prox}_{\nu g_1}$ and $\text{prox}_{\nu g_2}$ are very simple operations. We start by computing $\text{prox}_{\nu g_1}$. Let $\mathbf{Q} \in \mathbb{C}^{N \times N}$ be a symmetric matrix and suppose it has an eigenvalue decomposition $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$, where \mathbf{U} is the orthogonal matrix of eigenvectors and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$ is the diagonal matrix with the eigenvalues. Then, the proximity operator of νg_1 is computed as:

$$\text{prox}_{\nu g_1}(\mathbf{Q}) = \mathbf{U} \bar{\mathbf{\Lambda}}_\nu \mathbf{U}^\dagger, \quad (3.15)$$

where $\bar{\mathbf{\Lambda}}_\nu = \text{diag}((\lambda_1 - \nu)^+, \dots, (\lambda_N - \nu)^+)$ and $(a)^+ = \max(0, a)$ denotes the positive part of a . The operator $\bar{\mathbf{\Lambda}}_\nu$ performs a soft thresholding on the eigenvalues of \mathbf{Q} , to minimize

the nuclear norm, and also preserves only the positive eigenvalues, to project onto the set of positive-semidefinite matrices [57, 58]. The proximal operator of νg_2 is the projector onto the set of positive tensors in $\mathbb{R}^{N \times N \times N}$ which is computed by setting the imaginary part and the negative values of the real part of the input tensor to zero, i.e.

$$\text{prox}_{\nu g_2}(\mathcal{S}) = \{(\text{Re}(\mathcal{S}_{i,j,k}))^+\}_{1 \leq i,j,k \leq N}, \quad (3.16)$$

where $\text{Re}(\cdot)$ denotes the real part of a complex number [29].

The proximal operator of f_2 is the projector operator onto the set C_ϵ , which is computed as:

$$\text{prox}_{\nu f_2}(\mathcal{X}) = \mathcal{X} + \mathcal{T}_s^\dagger (\mathcal{P}_\epsilon (\mathcal{T}_s(\mathcal{X}) - \mathbf{y}_s) - \mathcal{T}_s(\mathcal{X}) + \mathbf{y}_s), \quad (3.17)$$

where $\mathcal{P}_\epsilon(\mathbf{r}) = \min(1, \epsilon/\|\mathbf{r}\|_2)\mathbf{r}$.

All the operations done in the computation of the proximal operators of f_1 and f_2 preserve tensor symmetry, provided that the symmetrized version \mathcal{T}_s of the measurement operator and a symmetrized data vector are used. These two are sufficient conditions to impose supersymmetry at each iteration of Algorithm 2, and consequently for the final tensor solution.

The memory requirement to solve this NM problem is dominated by the storage of the tensor, which is of size $\mathcal{O}(N^3)$. In terms of computation time, the algorithm is dominated at it each iteration by the application of the operator \mathcal{T}_s which computes N^2 2D FFTs of size N along each of the three dimensions, with an asymptotic complexity of $\mathcal{O}(N^3 \log N)$. These orders of magnitude obviously stand in stark contrast with those for the AM approach.

While the NM approach is much heavier than the AM approach in terms of memory requirements and computation complexity due to the drastically increased dimensionality of the unknown, the underlying convexity at the tensor level ensures essential properties of convergence to a global minimum of the objective function and independence to initialization, justifying a comparative analysis.

3.2.3.3 Nuclear minimization with sparsity

Following the lines of recent evolutions in radio interferometry [34, 35, 38] and in optical interferometry [32, 48], we decided to study the inclusion of a sparsity prior for the NM approach. As a first proof of concept we have chosen to promote the simplest sparsity – in image space – of the signal \mathbf{x} of interest, as this can be done simply through adopting a sparsity prior directly on the full tensor \mathcal{X} . While ℓ_0 -minimization would promote sparsity explicitly, we adopt the common convex relaxation relying on the ℓ_1 norm. Note that a nonweighted ℓ_1 norm is not a meaningful prior function as the tensor values are positive and sum up to unity. In that scenario, we resort to a reweighting scheme consisting in approaching both ℓ_0 -minimization on \mathcal{X} and rank minimization on $\mathcal{C}(\mathcal{X})$ by solving a

sequence of weighted ℓ_1 and nuclear norm minimization [24, 59], each of which is initialized with the solution of the previous problem (see sections 2.4.2 and 2.4.3).

The weighted- ℓ_1 and nuclear-norm minimization problem (NM-RW) thus reads as:

$$\min_{\mathcal{X} \in S} \|\mathcal{C}(\mathcal{X})\|_{*,w} + \lambda \|\mathcal{X}\|_{1,w} \quad \text{s. t.} \quad \|\mathbf{y}_s - \mathcal{T}_s(\mathcal{X})\|_2 \leq \epsilon,$$

where S denotes the same set as in (3.12) and $\|\cdot\|_{*,w}$ and $\|\cdot\|_{1,w}$ denote weighted nuclear and ℓ_1 norms respectively. Notice that the weights for the nuclear and the ℓ_1 norm are defined in a different form. In both weighted norms, each element of the vector to be reweighted should essentially be divided by its absolute value in the previous iteration. A stabilization parameter, δ , is necessary to define the weights properly, even when the signal value is zero. In the weighted ℓ_1 norm, each weight is defined as $w_{ijk} = \delta^{(t)} / (\delta^{(t)} + \mathcal{X}_{ijk}^{(t-1)})$, where t indicates the iteration of the reweighting process. λ is set to zero at the first iteration to avoid the use of a nonweighted ℓ_1 norm as a prior, as previously mentioned. In the following iterations, we heuristically set $\lambda^{(t)} = \alpha \|\mathcal{X}^{(t-1)}\|_\infty$, where $\|\cdot\|_\infty$ denotes the maximum absolute value of the tensor and $0 < \alpha < 1$ is a parameter to be tuned. In order to approximate the rank function through the weighted nuclear norm – i.e. the weighted ℓ_1 norm of the singular values $\sigma_i, i \in \mathbb{R}_+^N$ –, each weight is computed as the inverse of the singular value of $\mathcal{C}(\mathcal{X})$ at the previous iteration, $w_i = \delta^{(t)} / (\delta^{(t)} + \sigma_i^{(t-1)})$. The reweighting process stops when the relative variation between successive solutions $\|\mathcal{X}^{(t)} - \mathcal{X}^{(t-1)}\|_2 / \|\mathcal{X}^{(t-1)}\|_2$ is smaller than some bound or after the maximum number of iterations allowed is reached. Finally, the signal is extracted from the tensor using the rank-1 approximation algorithm [55], as mentioned in Section 3.2.3.1.

3.3 Simulations and results

In this section we evaluate the performance of the AM, NM and NM-RW algorithms through numerical simulations. Our optimization code^{||} was implemented in MATLAB and run on a standard 2.4 GHz Intel Xeon processor. Given the expected large memory requirements and long reconstructions time for the NM formulation, we consider small-size images with $N = 16^2 = 256$ for which the image vector occupies the order of 4 KB in double precision, while the size- N^3 tensor variable already takes the order of 100 MB. The memory requirement for the simple tensor variable would already rise to the order of 8 GB for a $32^2 = 1024$ image size.

For what the measurement setting is concerned, we assume random variable-density sampling in the 6D Fourier space, where low spatial frequencies are more likely to be sampled than high frequencies. In practice the sampling pattern is obtained by sampling *frequencies* independently along each of the 3 tensor dimensions from a bidimensional random Gaussian profile in the corresponding Fourier plane, associating the originally continuous

^{||}Code and test data are available at <https://github.com/basp-group/co-oi>.

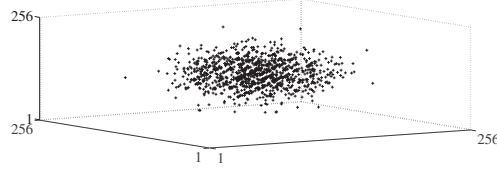


Figure 3.2: Example of variable-density sampling pattern in the discrete 6D Fourier space of \mathcal{X} of dimension N^3 , for a $N = 16^2$ image size and an undersampling regime of $M/N = 0.75$.

random points with the nearest discrete frequency. The sampling is carried out progressively, noting that if a product is sampled twice the result is discarded and repeating this procedure until M samples are obtained. Again this consideration only arises from the discrete setting adopted. Figure 3.2 presents a typical sampling pattern.

In all experiments we define the input signal-to-noise ratio as $\text{ISNR} = -10 \log(\sigma_n^2/e_y^2)$ where $e_y^2 = (1/M) \sum_a^M |y_a|^2$. The signal-to-noise ratio of a reconstruction $\bar{\mathbf{x}}$ is defined as $\text{SNR} = -10 \log(\|\bar{\mathbf{x}} - \mathbf{x}\|^2/\|\mathbf{x}\|^2)$. With this definition, the higher the SNR, the closer the recovered signal $\bar{\mathbf{x}}$ is from the original \mathbf{x} .

3.3.1 AM vs NM comparison

As a preliminary experiment, we provide a comparison of the performance of the NM approach defined in (3.12), with the equivalent minimization problem where the summation operator \mathcal{C} is replaced by the unfolding operator \mathcal{U} in the nuclear norm and where the positive-semidefiniteness constraint is discarded as it does not apply for non-square matrices. Both algorithms were tested on images constructed from 32 random spikes, with $\text{ISNR} = 30\text{dB}$. The positive spike values are taken uniformly at random and normalized to get unit flux, while positions are drawn at random from a Gaussian profile centered on the image. The graphs in Figure 3.3 represent the SNR and timing curves as a function of undersampling in the range $[0.25, 1]$. A total of 10 simulations per point are performed, varying the signal, as well as the sampling and noise realizations. Both approaches provide similar reconstruction qualities, with a smaller variability of the component summation approach, which is also slightly superior at low undersamplings. The component summation approach, running in the order of 10^3 seconds, is as expected significantly faster than the unfolding approach, running on average more than 10 times more slowly in the range $[0.5, 1]$. We therefore discard further consideration of the latter.

Having validated our NM approach in comparison with alternative state-of-the-art low tensor rank approaches, we compare its performance with that of the AM scheme. Firstly, we evaluate the reconstruction quality on images constructed from 32 and 64 randomly located spikes. The AM approach is also evaluated for varying reinitialization numbers: $n_{\text{ri}} \in \{1, 5, 10\}$. The graphs in Figure 3.4 represent the SNR curves as a function of undersampling in the range $[0.25, 1]$. A total of 50 and 10 simulations per point are performed for AM and NM respectively, varying the signal, as well as the sampling and

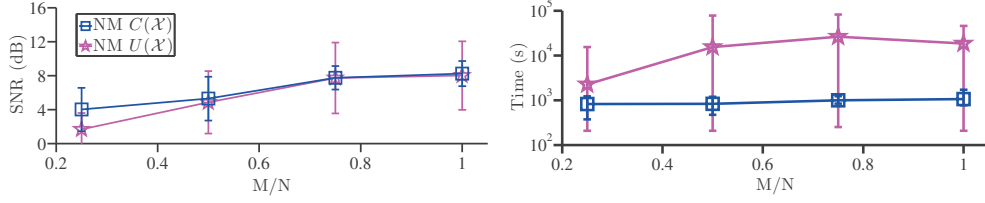


Figure 3.3: Reconstruction quality and timing comparison between the NM approach defined in (3.12), with the equivalent minimization problem where the summation operator \mathcal{C} is replaced by the unfolding operator \mathcal{U} . Tests done on $N = 16^2$ images with 32 randomly located spikes and $\text{ISNR} = 30\text{dB}$, for undersampling ratios M/N in the range $[0.25, 1]$. The SNR curves (left panel) represent average values over 10 simulations and corresponding 1-standard-deviation error bars. The timing curves (right panel) represent average values over 10 simulations and min-max error bars.

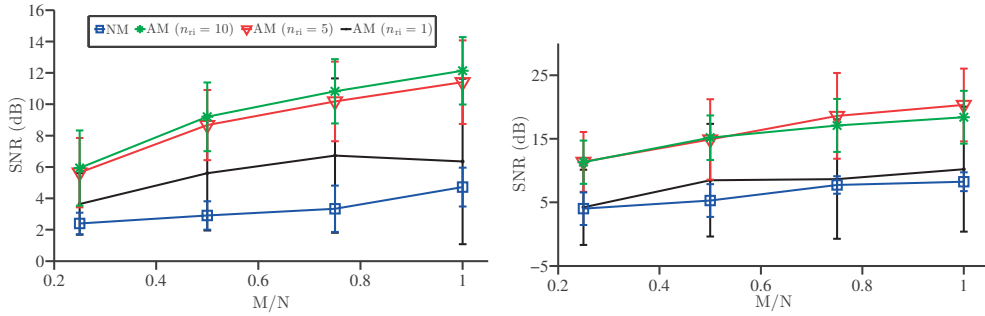


Figure 3.4: Reconstruction quality results for synthetic images of size $N = 16^2$ with randomly distributed spikes and $\text{ISNR} = 30\text{dB}$ for undersampling ratios M/N in the range $[0.25, 1]$. Left panel: 64 spikes. Right panel: 32 spikes. The curves represent the average SNR values over multiple simulations (50 for AM and 10 for NM) and corresponding 1-standard-deviation error bars.

noise realizations. The results show a clear superiority of AM relative to NM in terms of average reconstruction quality. Both approaches exhibit nonnegligible variability. The dependency of the nonconvex AM approach to initialization is clearly illustrated by the $n_{\text{ri}} = 1$ and $n_{\text{ri}} = 5$ curves, confirming the importance of the multiple reinitializations. We also observe a saturation between $n_{\text{ri}} = 5$ and $n_{\text{ri}} = 10$. As expected from asymptotic complexity considerations, AM runs significantly faster than NM, with reconstructions in the order of 10^2 seconds for $n_{\text{ri}} = 5$, approximately 10 times faster than NM.

Secondly, simulations are performed in an identical setting on realistic images representing low-resolution versions of the Eta Carinae star system, of a simulated rapidly rotating star, and of the M51 Galaxy**. The multiple simulations per point are performed by varying the sampling and noise realizations. The graphs in Figures 3.5, 3.6, and 3.7 present the SNR curves as a function of undersampling in the range $[0.25, 1]$ (AM only reported for $n_{\text{ri}} = 5$), confirming the previous results on random images. Reconstructed

**Images from [48], downloaded from the JMMC service at apps.jmmc.fr/oidata/shared/srenard/.

images are also reported, providing visual confirmation of the superiority of AM relative to NM over the full undersampling range. In both approaches, the visual quality difference between the reconstructions with, respectively, best and median SNR values illustrates the variability of the reconstruction quality. The NM approach suffers from a significantly larger visual degradation of median SNR value at $M = 0.25N$ than AM. This degradation appears at larger sampling ratios for M51.

Let us highlight that, while only 5 reinitializations are necessary in the AM approach in low dimension to reach saturation, additional experimental tests on random signals of size $N = 64^2$ show that $n_{\text{ri}} = 20$ or larger is necessary for a meaningful reconstruction, thereby emphasizing the convergence problem due to nonconvexity in higher dimension. Also, computation time scales linearly with n_{ri} and can rapidly blow up in this context.

3.3.2 NM vs NM-RW comparison

To compare the performance of the linear NM-RW and NM approaches, we first evaluate the reconstruction quality on sparse images made of 8 and 16 randomly located spikes. The SNR curves in Figure 3.8 are built from 10 simulations per point, varying the signal, the sampling and noise realizations. The results show a clear improvement on the SNR when accounting for sparsity. In Figure 3.9, the effect of the reweighting scheme can be appreciated on an illustration representing a sparsified version of Galaxy M51^{**}. Reweighted images (second row) are less blurred and their support is clearly better defined.

3.4 Conclusion

We have proposed a novel linear formulation of the optical-interferometric imaging problem in terms of the supersymmetric rank-1 order-3 tensor formed by the tensor product of the vector representing the image sought with itself. In this context, we proposed a linear convex approach for tensor recovery with built-in supersymmetry, and regularizing the inverse problem through nuclear norm minimization. We have also studied a non-linear nonconvex alternate minimization approach where supersymmetry is relaxed while the rank-1 constraint is built-in. While the former approach is associated with drastically increased dimensionality of the unknown, the underlying convexity ensures essential properties of convergence to a global minimum of the objective function and independence to initialization, justifying its analysis. Simulation results in low dimension show that the AM scheme provides significantly superior imaging quality than the NM approach, in addition to be much lighter in its memory requirements and computation complexity. Another set of results in higher dimension however suggests that the number of necessary reinitializations for the nonconvex AM scheme rapidly increases with N . This state of things clearly calls for further considerations of a purely convex approach. As a first step in this direction, we have studied the inclusion of a sparsity prior in the convex formulation

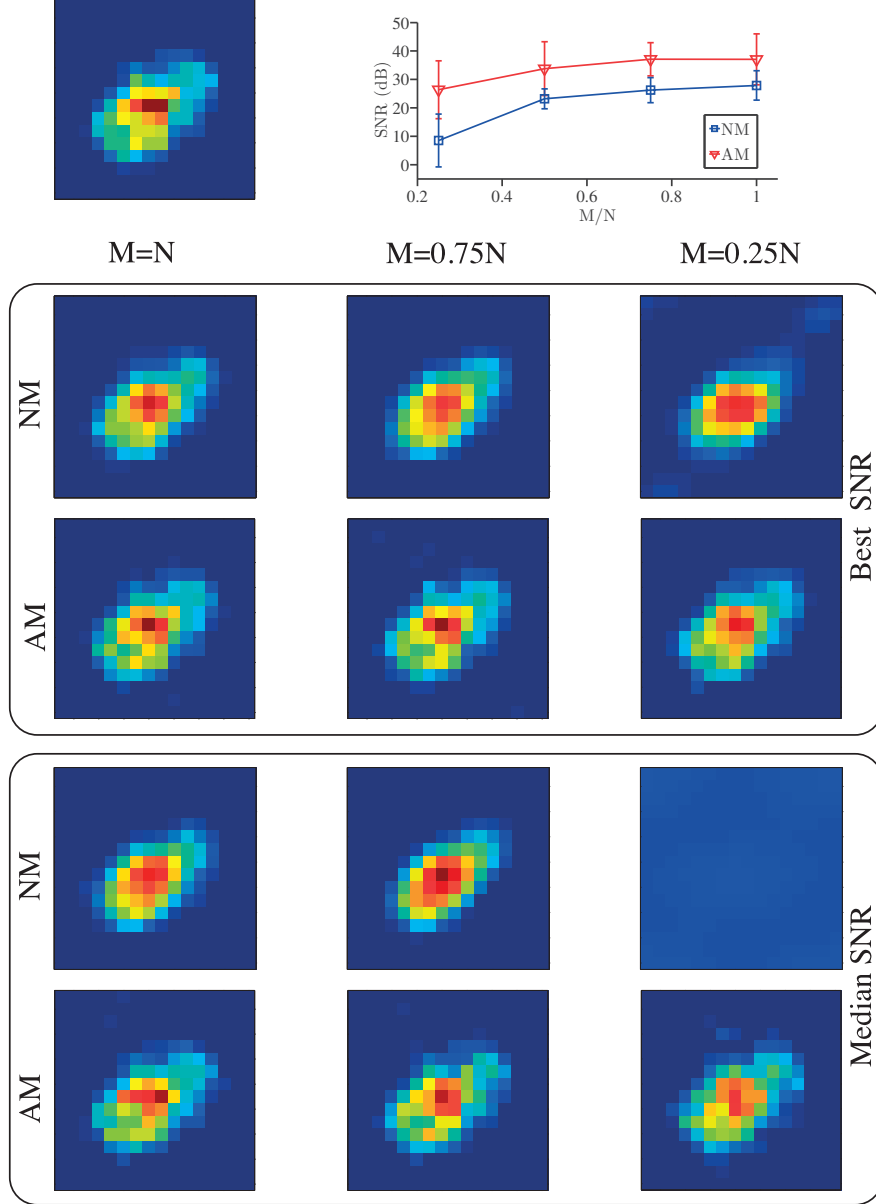


Figure 3.5: Eta Carinae star system illustration ($N = 16^2$, $\text{ISNR} = 30\text{dB}$). Top row: original image and SNR graph. The curves represent the average SNR values over multiple simulations (50 for AM and 10 for NM) and corresponding 1-standard-deviation error bars. Second and third rows: NM (second) and AM for $n_{\text{ri}} = 5$ (third) reconstructions with best SNR for $M = N$ (left), $M = 0.75N$ (center) and $M = 0.25N$ (right). Fourth and bottom rows: NM (fourth) and AM for $n_{\text{ri}} = 5$ (bottom) reconstructions with median SNR for $M = N$ (left), $M = 0.75N$ (center) and $M = 0.25N$ (right).

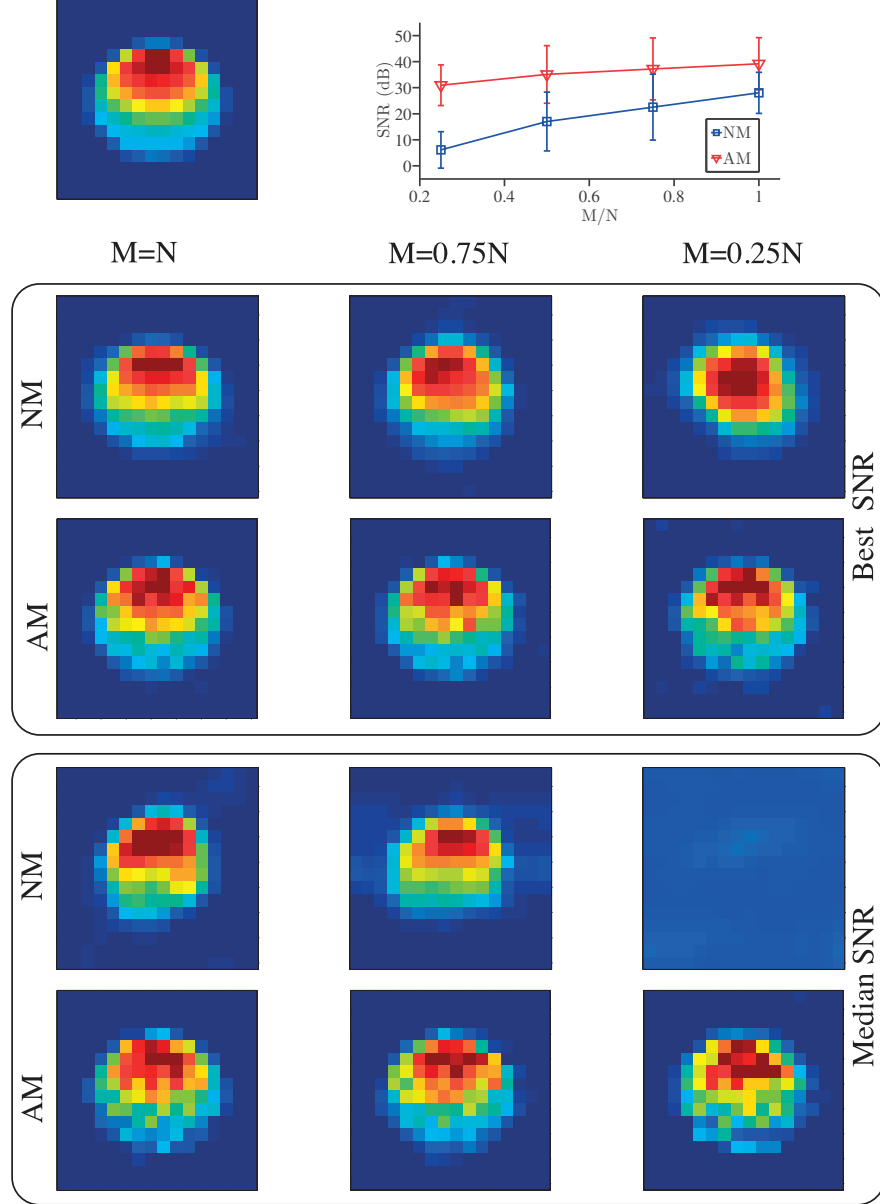


Figure 3.6: Rapidly rotating star illustration ($N = 16^2$, $\text{ISNR} = 30\text{dB}$). Top row: original image and SNR graph. The curves represent the average SNR values over multiple simulations (50 for AM and 10 for NM) and corresponding 1-standard-deviation error bars. Second and third rows: NM (second) and AM for $n_{\text{ri}} = 5$ (third) reconstructions with best SNR for $M = N$ (left), $M = 0.75N$ (center) and $M = 0.25N$ (right). Fourth and bottom rows: NM (fourth) and AM for $n_{\text{ri}} = 5$ (bottom) reconstructions with median SNR for $M = N$ (left), $M = 0.75N$ (center) and $M = 0.25N$ (right).

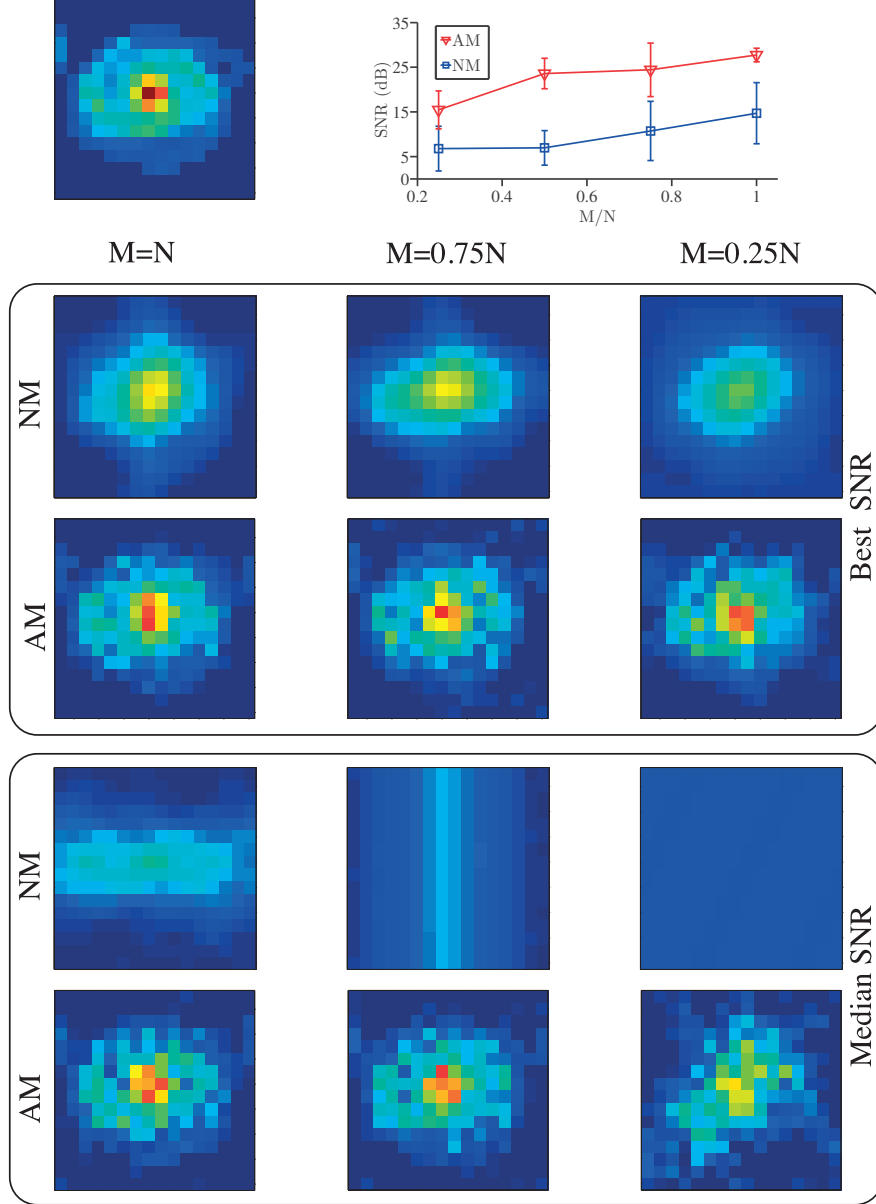


Figure 3.7: M51 Galaxy illustration ($N = 16^2$, $\text{ISNR} = 30\text{dB}$). Top row: original image and SNR graph. The curves represent the average SNR values over multiple simulations (50 for AM and 10 for NM) and corresponding 1-standard-deviation error bars. Second and third rows: NM (second) and AM for $n_{\text{ri}} = 5$ (third) reconstructions with best SNR for $M = N$ (left), $M = 0.75N$ (center) and $M = 0.25N$ (right). Fourth and bottom rows: NM (fourth) and AM for $n_{\text{ri}} = 5$ (bottom) reconstructions with median SNR for $M = N$ (left), $M = 0.75N$ (center) and $M = 0.25N$ (right).

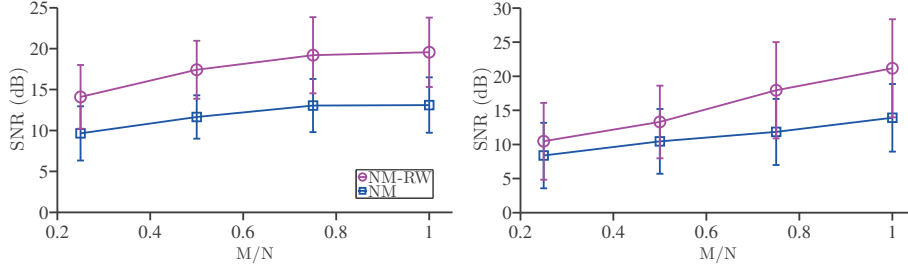


Figure 3.8: Reconstruction quality results for synthetic images of size $N = 16^2$ with randomly distributed spikes and $\text{ISNR} = 30\text{dB}$ for undersampling ratios M/N in the range $[0.25, 1]$. Left panel: 8 spikes. Right panel: 16 spikes. The curves represent the average SNR values over 10 simulations and corresponding 1-standard-deviation error bars.

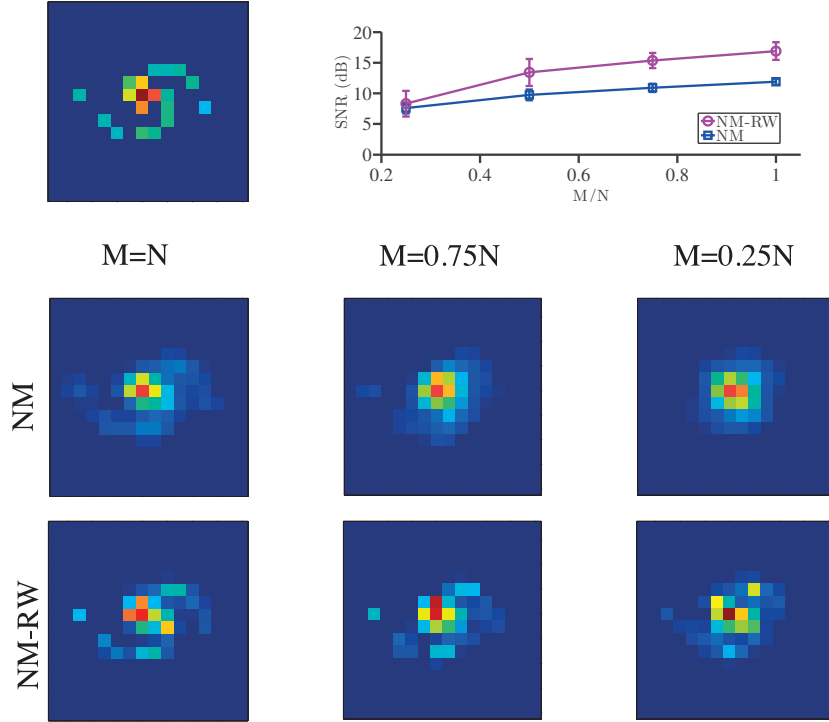


Figure 3.9: Sparsified version of M51 Galaxy illustration ($N = 16^2$, $\text{ISNR} = 30\text{dB}$). Top row: original image and SNR graph. The curves represent the average SNR values over 10 simulations and corresponding 1-standard-deviation error bars. Second and third rows: NM (second) and NM-RW (bottom) reconstructions with best SNR for $M = N$ (left), $M = 0.75N$ (center) and $M = 0.25N$ (right).

and, in that setting, numerical simulations confirm a clear improvement in the quality of the reconstruction of sparse images. We point out though, that in the present investigations only the simplest case of sparsity – sparsity in image space – has been considered, as a proof of concept. Future work should investigate the effects of assuming different kinds

of sparsity priors, as suggested in recent approaches for radio interferometry [60].

Our approaches should also be studied in a more realistic setting with exact power spectrum and bispectrum measurements in the continuous domain and for different noise statistics, and explicitly compared to existing MiRA and WISARD implementations. The linear approaches NM and NM-RW are extremely exacting from a computational standpoint so that software and hardware optimization should also be studied to solve the problem for higher dimension images, e.g. using graphics processing units [61]. Recent results studying the uniqueness of the solution of the phase retrieval problem for sparse signals are presented in [62]. Further research should also analyze our results and formulation in full view of this new theoretical framework.

The work presented in this chapter has been published in [3, 4]. We highlight that in [63], the authors enhance the present work by further study of the AM approach. They propose an algorithm that accounts for sparsity and presents convergence guarantees, missing in our original formulation. Numerical simulations in [63] show that the inclusion of a sparsity prior improves significantly the quality of their reconstructions, analogously to what we observe for the NM approach. Furthermore, they successfully extend the current model to hyperspectral imaging and demonstrate the superiority of their global approach compared to single-channel reconstruction.

Part II

Diffusion Magnetic Resonance Imaging

Background on diffusion MRI

4.1 Introduction

Diffusion Magnetic Resonance Imaging (dMRI) is sensitive to the Brownian motion of water molecules, i.e. its random displacement in a fluid due to thermal energy. This erratic movement is described in statistical terms by a displacement distribution indicating the proportion of molecules that have been displaced a specific distance in a specific direction. Typically, the displacement distribution for free water is a Gaussian function. However, in organs with ordered structures, such as the brain, water does not diffuse equally in all directions [64]. This phenomenon, known as *anisotropic* diffusion, is exploited to study the structure of spatial order in living organs in a non-invasive way.

This chapter aims to familiarize the reader with the principles of dMRI applied to brain imaging and it does not contain any original contribution. The chapter is organized as follows: Section 4.2 explains the principles behind this imaging modality and how the diffusion process is characterized. Section 4.3 describes how and why dMRI can be used in brain imaging to recover the properties of its fiber bundles (i.e. axon tracts), from both neuroscientific and clinical perspectives. Sections 4.4 and 4.5 review the main state-of-the-art methods for fiber reconstruction (orientation and microstructure recovery, respectively). The take-home messages of this chapter are summarized in Section 4.6.

4.2 Principles of diffusion Magnetic Resonance Imaging

Thanks to Stejskal and Tanner’s experiments and proofs [65], a methodology was defined to measure diffusion with MRI. They introduced the Pulse Gradient Spin Echo (PGSE) sequence that makes MR imaging sequences sensitive to diffusion by inserting two additional magnetic field gradient pulses subsequently besides the standard ones used for spatial encoding. This protocol allows a clear distinction between the encoding time δ (pulse duration) and the diffusion time Δ (time between the two gradient pulses). After applying the first gradient, protons’ phase changes along the direction of the gradient. When then, a second gradient of opposite magnitude (with the same direction and time period) is applied, the phase-shift induced by the first gradient can be reversed. If the

protons moved in between the application of the gradients the net phase accumulation cannot be exactly zero. As a result, in each voxel, we get a distribution of phase-shifts which results in a loss of coherence and therefore to a decrease of the signal amplitude (Figure 4.1). The wider the spread of displacements, the larger the signal damping. Thus, by applying a pair of gradient pulses before the data acquisition we make the resultant image sensitive to motional processes, such as diffusion.

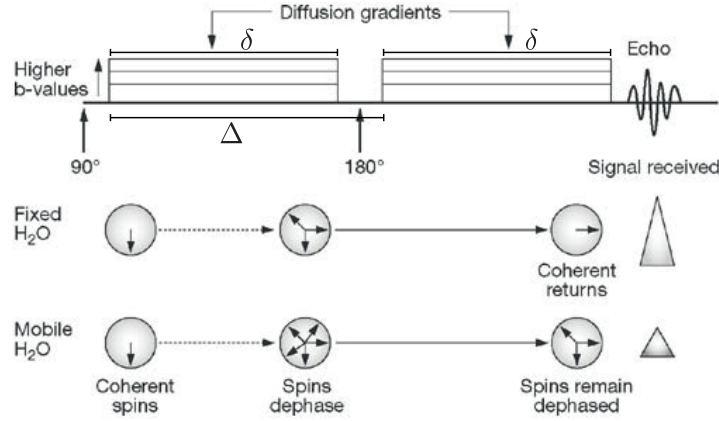


Figure 4.1: Diagram of a PGSE sequence and visual understanding of signal loss due to the dephasing of spins. *Part of the illustration is taken from [66]. Image used with permission.*

Each diffusion gradient is represented as a 3D vector \mathbf{q} , oriented in its direction and with a magnitude q proportional to its strength. The new 3D-space determined by the coordinates of \mathbf{q} is commonly named after q -space. The b -value, most commonly used when characterizing the gradient, is proportional to the product of the square of the gradient strength q and the diffusion time interval ($b \propto q^2 \cdot \Delta$). A single application of the PGSE sequence produces one brain image with a given diffusion weighting that corresponds to one point of the q -space. The resulting has already some diffusion information: where the main diffusion direction is aligned with the applied diffusion gradient, the intensity of the signal is markedly decreased, and the region therefore appears darker on the image (see figure 4.4 A). This is the principle behind the simplest diffusion imaging technique, known as Diffusion Weighted Imaging (DWI). Multiple repetitions of the sequence, each with a different diffusion weighting, are necessary if we want to characterize the entire diffusion process.

The diffusion process is represented by a 3D probability density function (pdf) of diffusion displacements at every point, i.e. for each voxel. This function $p(\mathbf{r}) \in \mathbb{R}^3$ is known as the Ensemble Average Propagator (EAP) for every \mathbf{r} coordinate vector in real

space and it relates to the diffusion signal $S(\mathbf{q})$ through a 3D Fourier transform [67], as follows:

$$\text{EAP} \equiv \mathbf{p}(\mathbf{r}) = \int_{\mathbb{R}^3} d\mathbf{q} S(\mathbf{q}) e^{-2\pi i \mathbf{q} \cdot \mathbf{r}}. \quad (4.1)$$

Equation (4.1) defines a relationship between the propagator (or real) space, with coordinates \mathbf{r} , and the signal (or Fourier) space, the q -space, through a Fourier transform.

4.3 dMRI in the brain

dMRI can provide very useful information about the organization of the tissue in the brain. In fact, the grey matter (GM) does not have an ordered fiber structure. Therefore, GM and the cerebrospinal fluid (CSF), that can be found in the ventricles, are typically isotropic media. However, high *anisotropy* can be observed in the white matter (WM), made of axons also called tracts. Experimental evidences point at the cell membrane as the main tissue component responsible for this *anisotropy* of molecular diffusion in the WM (see Figure 4.2). The degree of myelination of the individual axons and the density of cellular packing seem to merely modulate the anisotropy as measured with dMRI, with microtubules and neurofilaments playing only a minor role [64].

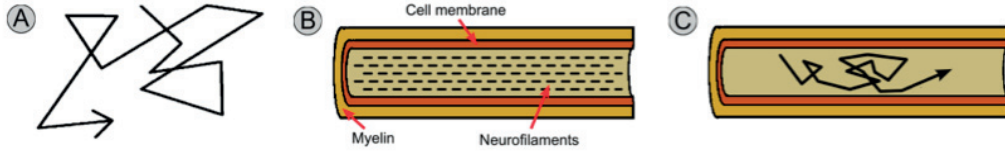


Figure 4.2: (A) Random motion of water molecules (free diffusion). (B) Water diffusion in the brain is mainly restricted by the membranes of the axons. (C) Inside the axons, water diffuses mainly along the axis of the axon. *Part of the illustration is taken from [68]. Image used with permission.*

Assessing the direction(s) with maximal diffusion remains one of the main interests of the study of the WM using dMRI. Thus, the Orientation Distribution Function (ODF) is defined to integrate the angular content of the EAP (4.1). The ODF gathers this angular information through radial integration, as follows:

$$\text{ODF} \equiv \mathbf{o}(\hat{r}) = \int_{\mathbb{R}_+} dr r^2 p(r, \hat{r}), \quad (4.2)$$

where r and \hat{r} correspond to the angular and radial components of \mathbf{r} . An ODF can be computed in every voxel, providing a better visualization of the diffusion at that position. The ODF can also be interpreted as a convolution on the sphere of a fiber response function or kernel with a Fiber Orientation Distribution function (FOD). The FOD is a real-valued

function on the unit sphere (\mathbb{S}^2) that indicates the orientation and the volume fractions of the fiber populations in a voxel (see Figure 4.3).

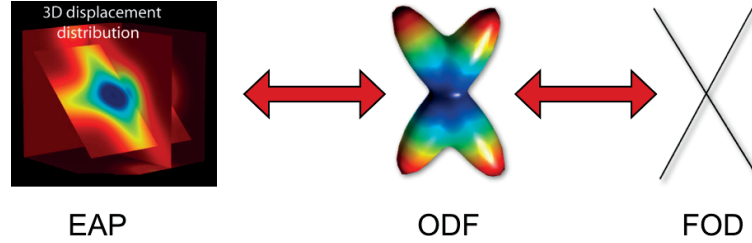


Figure 4.3: The propagator EAP (left) is a 3D pdf that indicates the probability of water displacements. The ODF (middle) is a function on \mathbb{S}^2 that represents the probability of diffusion along a given direction. The FOD (right), also a function on \mathbb{S}^2 , can be interpreted as a probability of having a fiber along a given direction. *Part of the illustration is taken from [68]. Image used with permission.*

The structural neuronal connectivity of the brain can be mapped in a non-invasive way thanks to the anisotropy of diffusion in the WM. The study of this connectivity is of major importance in a fundamental neuroscience perspective – for developing our understanding of the brain – but also in a clinical perspective, with particular applications for the understanding of stroke, schizophrenia, or Parkinson’s disease. Note that a precise mapping of the connectivity is commonly performed by so-called fiber-tracking or tractography algorithms. These algorithms produce trajectories capturing coherent orientations of maximal diffusion that are likely to represent real axonal fibers*. Therefore they heavily rely on the quality of the fiber orientation recovery in each voxel.

More recently, the anisotropy of diffusion in the brain has started to be exploited to characterize the microstructural properties of the WM, like axon diameter and density. These features are related to conduction velocity [69] and play an important role in the performance of the WM tracts. Also, the microstructure organization of specific areas of the brain changes in subjects affected by certain pathologies, such as multiple sclerosis [70]. As a consequence, assessing both the orientation of the WM tracts and their microstructure characteristics through diffusion MR imaging *in vivo* represents today a fundamental tool for neuroscience as well as from the clinical point of view.

A great variety of approaches have been proposed to tackle the problem of intra-voxel fiber estimation from dMR measurements. In what follows, we cover in detail the main state-of-the-art dMR imaging techniques so that their strengths as well as their limitations can be understood. We distinguish two groups of local reconstruction techniques: those mainly concerned to recover the *orientation* of the fiber bundles in each voxel (section 4.4) and those that go a step further and aim to recover the *microstructure* configuration of

*Of course, there are several orders of magnitude between the resolution of the MR acquisitions and the diameter of the axons. Therefore, a single reconstructed trajectory has to be thought of as representative of a huge coherent set of real anatomical fibers.

the fibers (section 4.5).

4.4 Fiber orientation reconstruction

Many methods have been proposed to assess the orientation of the main fiber bundles voxelwise. Hereafter we present in detail the three of them that appear more relevant in the context of this thesis: Diffusion Tensor Imaging, Diffusion Spectrum Imaging and Spherical Deconvolution methods. We name and refer to others in section 5.1.

4.4.1 Diffusion Tensor Imaging, DTI

The Diffusion Tensor Imaging (DTI) technique was one of the first approaches proposed to provide a unified description of the diffusion process from a series of DWI images. It was introduced by Bassar in 1994 [71]. The DT model is based on the hypothesis that the diffusion follows a Gaussian damping with \mathbf{q} , or equivalently, exponential in b . In an isotropic medium, the attenuation of the MRI signal can be described as $A = e^{-bD}$, where D , the diffusion coefficient, is a scalar. However, a tensor \mathcal{D} is required to describe the diffusion process along the three directions of our reference frame when it takes place in an anisotropic medium. To fully determine the diffusion tensor, diffusion-weighted images along several gradient directions must be collected. As the diffusion tensor is symmetric, measurements along only six directions are mandatory -instead of nine- (Figure 4.4 B), along with an image acquired without diffusion weighting ($b = 0$). Once the set of DW images is acquired, linear regression techniques can be used to estimate the full tensor \mathcal{D} . Its largest eigenvalue and corresponding eigenvector describe the intensity and the principal direction of diffusion, which are then associated with the orientation of the underlying fiber bundle.

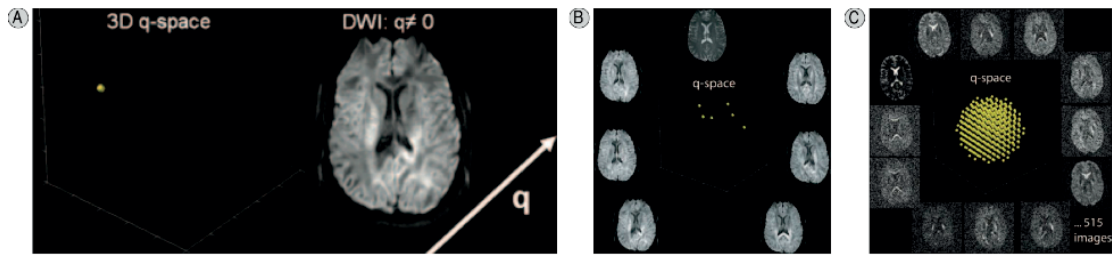


Figure 4.4: (A) Illustration of a DWI image from sampling the signal at a single \mathbf{q} -point. (B) In DTI, the q -space is sampled at least at 6 \mathbf{q} -points. (C) In DSI, the q -space is densely sampled, each \mathbf{q} -point corresponding to a different gradient orientation and strength. *Part of the illustration is taken from [68]. Image used with permission.*

The diffusion tensor is normally visualized as an ellipsoid with the principal axes along the eigenvectors of \mathcal{D} , and with the length of these axes proportional to the corresponding eigenvalues. These eigenvalues can be considered as unidimensional diffusion coefficients,

or *diffusivities*, in the main directions of diffusion in the medium, so for instance, the eccentricity of the ellipsoid gives us information about the degree of anisotropy.

The Tensor Model assumes gaussian diffusion and gaussian functions have only one maximum. In voxels with presence of multiple fiber orientations (fiber crossings or branching), DTI will recover a single fiber orientation, corresponding to the mean of the true underlying directions. Consequently, the DT model is not valid in regions of the brain where there are fiber crossings. In such areas, other methods providing a higher angular resolution are needed.

4.4.2 Diffusion Spectrum Imaging, DSI

DSI is a model-free imaging technique that samples densely the q -space from which the displacement distribution can be later recovered directly exploiting the Fourier relationship described in (4.1). It therefore requires the acquisition of *many* DW images, each of them corresponding to a different \mathbf{q} -point distributed in a Cartesian grid (Figure 4.4).

With this imaging technique, fiber orientations are also associated with directions of maximum diffusion, but since there is no restriction on one single diffusion direction, it appears as a more suitable approach to detect fiber crossings. DSI provides very good angular resolution when sampling densely the q -space. Nevertheless, the main drawback of DSI strongly relates to its complexity. Measuring the complete 3-dimensional diffusion function requires long acquisition times and thus it is not appropriate for all clinical applications.

4.4.3 Methods based on spherical deconvolution

Spherical deconvolution (SD) methods [5, 72, 73] reformulate the imaging problem from dMR data as a deconvolution problem on the sphere under the following two assumptions:

- (i) There is no exchange of water between different fiber bundles over the time of a dMR measurement, meaning that the signal attenuation in a voxel with the presence of N fiber populations can be expressed as $S = \sum_N f_i S_i$, where f_i is the volume fraction of the i th bundle and S_i its corresponding signal damping.
- (ii) All fiber bundles in the brain share the same diffusion attenuation profile, i.e. if there is any difference among them it must be due to a partial-volume effect.

Under conditions (i) and (ii), the diffusion signal in every voxel can be expressed as a convolution over the unit sphere of a response function and the FOD:

$$S(\mathbf{q}) = S_0 \int_{\mathbb{S}^2} K(\mathbf{q}, \hat{\mathbf{u}}) f(\hat{\mathbf{u}}) d\hat{\mathbf{u}}. \quad (4.3)$$

In (4.3), S_0 corresponds to the signal without diffusion weighting, K represents the kernel or response function, f stands for the FOD and $\hat{\mathbf{u}} \in \mathbb{S}^2$ is a unitary vector. Assuming

this kernel K can be estimated a priori, the measurement process can be expressed as a system of linear equations:

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\eta}, \quad (4.4)$$

where \mathbf{x} corresponds to the vector of coefficients of the FOD to be estimated, \mathbf{y} represents the vector of measurements, the so-called dictionary Φ models the convolution operator and $\boldsymbol{\eta}$ accounts for the acquisition noise. Several approaches have been proposed to solve this ill-posed problem and the quality of the reconstructions as well as the minimum number of \mathbf{q} -points that needs to be sampled depend a lot on the regularization strategy. The reader will find further discussion on the topic in chapter 5.

4.5 Microstructure imaging techniques

Most microstructure imaging techniques describe the brain tissue using multi-compartment models distinguishing, for instance, axons, glial cells and extra-axonal space. Subsequently, the microstructure properties are recovered by modeling the signal decay in each of the considered tissue compartments and assuming the measured signal as a combination of all of them (see Figure 4.5).

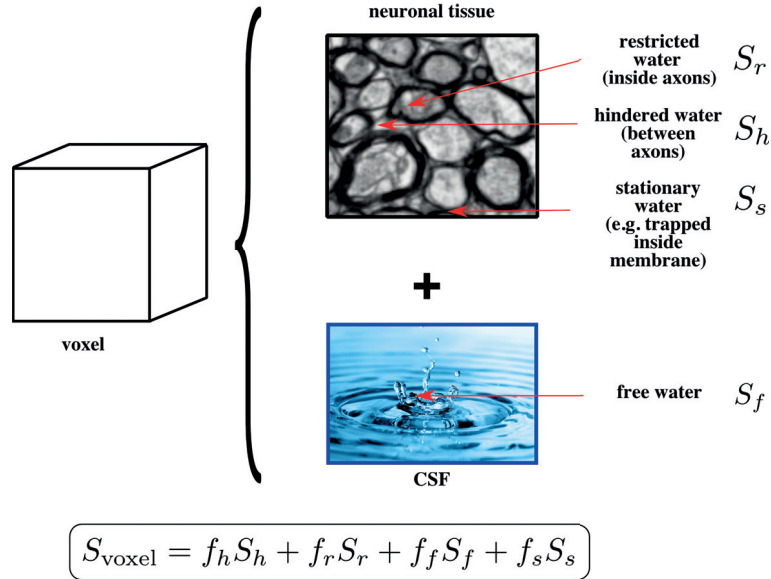


Figure 4.5: Schematic representation of the tissue multi-compartment model for ActiveAx [1]. In every voxel, the signal decay S_{voxel} is expressed as the weighted sum of the contributions of each compartment. f_r, f_h, f_f and f_s represent the volume fractions corresponding to the restricted, hindered, free and stationary compartments, respectively. *Part of the illustration is taken from [74]. Image used with permission.*

The main differences between different microstructure imaging methods depend on the number of tissue compartments they consider, the chosen model for the diffusion

in each compartment and whether they assume water exchange between them or not. The reader can find an exhaustive survey on many of the state-of-the-art microstructure imaging techniques in [75]. We have summarized their principal characteristics in Table 4.6, indicating for each image modality the signal model used in each compartment and the features it estimates. In Table 4.6, we borrow Panagiotaki’s taxonomy [75] to refer to the diverse *signal models*. We briefly describe them hereafter and we refer the reader to [75] for further study:

- **Intra-axonal** refers to the signal decay due to water diffusion inside the axons. Subsequently we also refer to it as *restricted compartment*.
 - ★ *stick*: provides the signal decay that would correspond to diffusion inside an idealized 0-radius cylinder.
 - ★ *cylinder*: indicates the signal corresponding to diffusion inside a cylinder.
 - ★ *GDRCylinders*: indicates the signal decay corresponding to diffusion inside a group of cylinders whose radii are drawn from a Γ -distribution.
- **Extra-axonal** refers to the signal decay due to water diffusion outside and between axons. Subsequently we also refer to it as *hindered compartment*. The different models express hindered diffusion as gaussian tensors, with different degrees of freedom.
 - ★ *ball*: corresponds to isotropic diffusion with a single diffusivity parameter.
 - ★ *zeppelin*: stands for an anisotropic cylindrically-symmetric tensor with two diffusivity parameters: d_{\parallel} , parallel to the principal direction of the tensor and d_{\perp} , perpendicular to it.
 - ★ *tensor*: stands for an anisotropic tensor with three diffusivity parameters: d_{\parallel} , parallel to the principal direction of the tensor and d_{\perp_1}, d_{\perp_2} , perpendicular to it.
- **Stationary** refers to the signal decay due to water trapped in other cellular structures, such as glial cells.
 - ★ *sphere*: models the water diffusion inside bodies with spherical boundaries.
 - ★ *dot*: refers to a 0-radius sphere and is intend to model water that does not move.
- **Free** refers to the signal decay due to free water in a non-restrictive environment, such as CSF.

The *microstructure parameters* that can be estimated with each technique depend on its chosen tissue/signal model. When interpreting Table 4.6, consider that a specific

feature is estimated by a particular technique if a green tick appears in the corresponding row. A red cross indicates that the referred technique does not estimate that particular parameter, either because it is meaningless according to the chosen signal model or because it is fixed *a priori*. In general they can all be related either to the orientation, the size or the packing density of the axons; hereafter we list those mentioned on Table 4.6:

- Fiber bundle main **orientation**.
- **Volume fractions** corresponding to every considered compartment.
- **Diffusivities**: each technique chooses to fix some (or all) of the diffusivity parameters to biologically-plausible values characteristic to the medium and to estimate the rest from the data[†]. In most cases, they are related to the intra-axonal volume fraction through a tortuosity model [76].
- Axon Diameter Distribution (**ADD**).
- Axon **Diameter Index**: It is a single summary statistics that corresponds to the diameter that would produce the signal decay best matching the average signal decay over the distribution. Since large axons contain more water than the small ones, they also contribute “more” to the diffusion signal. Therefore, this diameter index correlates to a *weighted mean* axonal diameter, in which the contribution of each axon to the total signal is proportional to the square of its diameter [1].
- Orientation **dispersion** around the mean fiber orientation.

All microstructure imaging techniques mentioned in Table 4.6 require computationally very expensive nonlinear procedures to fit their models to the data [9]. Moreover, they are only valid in regions with one single fiber population, which makes them inappropriate to characterize the microstructure of the majority of voxels in the brain. In the next subsection, we present a general framework in which any of the existing techniques can potentially be reformulated into a linear problem that can be easily and very rapidly solved using convex optimization.

4.5.1 The AMICO framework

In [9], the authors define a flexible framework for microstructure imaging named after AMICO, Accelerated Microstructure Imaging via Convex Optimization. The authors get inspired by spherical deconvolution methods that formulate the problem of fiber orientation recovery in a voxel as a linear inverse problem, provided the response function of a

[†]In the table we do not specify which are the fixed/estimated diffusivities for every model since it appears out of the scope of this thesis. A red cross indicates that the method does not estimate any diffusivity and a green tick, some of them. The reader can refer to the original papers for more details.

		SIGNAL MODELING				ESTIMATED PARAMETERS					
		INTRA-AXONAL	EXTRA-AXONAL	CSF	STATIONARY WATER (e.g. glial cells)	ORIENTATION	VOLUME FRACTIONS	DIFFUSIVITIES	ADD	DIAMETER INDEX	DISPERSION
Non-permeable compartments	Stick and Ball	stick	ball	x	x	✓	✓	✗	✗	✗	✗
	CHARMED	GDRCyinders	tensor	x	x	✓	✓	✓	✗	✗	✗
	AxCaliber	GDRCyinders	tensor	x	x	✗	✓	✓	✓	✗	✗
	extended AxCaliber	GDRCyinders	tensor	ball	x	✗	✓	✓	✓	✗	✗
	ActiveAx	Cylinder	zeppelin	ball	dot	✓	✓	✓	✗	✓	✗
	NODDI	set of sticks with an orientation distribution	zeppelin	ball	x	✓	✓	✗	✗	✗	✓
Permeable compartments	Stanisz	prolate ellipsoids	zeppelin	x	sphere	✓	✓	✓	✗	✓	✗

Figure 4.6: Summary-table of the main features of the following microstructure imaging techniques: Stick and Ball [77], CHARMED [78, 79], AxCaliber [80], extended AxCaliber [74], ActiveAx [1, 81], NODDI [82], Stanisz [83].

single fiber can be estimated (see section 4.4.3). To extend this idea to microstructure imaging, in AMICO the original problem is decoupled into two independent subproblems:

- (i) estimation the main orientation ($\boldsymbol{\mu} \in \mathbb{S}^2$) of the fiber population under study, and
- (ii) estimation of its main microstructure features.

Assuming the presence of a single fiber population in the voxel, step (i) is easily performed using DTI [71]. Once $\boldsymbol{\mu}$ is known, the microstructure mapping problem is expressed in terms of a linear formulation, as follows:

$$\mathbf{y} = \Phi_{\boldsymbol{\mu}} \mathbf{x} + \boldsymbol{\eta}, \quad (4.5)$$

being $\mathbf{y} \in \mathbb{R}_+^m$ the vector of diffusion measurements, $\mathbf{x} \in \mathbb{R}_+^n$ the coefficients to be estimated and $\boldsymbol{\eta}$ the acquisition noise. In (4.5), $\Phi_{\boldsymbol{\mu}}$ is a linear operator or dictionary that accounts for the signal decay from different compartments (oriented along direction $\boldsymbol{\mu}$) and therefore, it is designed according to the imaging modality that is meant to reformulate. In [9], the authors demonstrate the linearization of ActiveAx [1] and NODDI [82], even if the AMICO framework can be also applied to other microstructure imaging models. Hereafter, we exemplify this linearization process by detailing the construction of the dictionary for ActiveAx.

ActiveAx as a linear system is built from different sub-matrices:

$$\Phi_{\boldsymbol{\mu}} = [\Phi_{\boldsymbol{\mu}}^r | \Phi_{\boldsymbol{\mu}}^h | \Phi_{\boldsymbol{\mu}}^i]. \quad (4.6)$$

In equation (6.6), sub-matrices $\Phi_\mu^r \in \mathbb{R}^{m \times N_r}$, $\Phi_\mu^h \in \mathbb{R}^{m \times N_h}$ and $\Phi_\mu^i \in \mathbb{R}^m$ model, respectively, the intra-axonal, extra-axonal and isotropic contributions to the diffusion signal along the direction μ . Each atom in sub-matrices Φ_μ^r models the diffusion signal corresponding to water molecules *restricted* within parallel cylinders of a specific diameter. Alternatively, the atoms in sub-matrices Φ_μ^h describe the *hindered* space between the axons and a single atom in Φ_μ^i accounts for any isotropic contribution. N_r and N_h represent, respectively, the number of different axon radii and hindered environments considered to build the dictionary ($n = N_r + N_h + 1$).

Equation (4.5) is then solved as a Tikhonov-regularized least-squares problem as follows:

$$\min_{\mathbf{x} \geq 0} \frac{1}{2} \|\Phi_\mu \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \frac{1}{2} \|\mathbf{x}\|_2^2, \quad (4.7)$$

where $\|\cdot\|_2$ is the standard ℓ_2 norm and parameter $\lambda > 0$ controls the trade-off between data and regularization terms. Equation (4.7) can be solved using fast convex optimization methods and therefore, AMICO provides *an acceleration factor of several orders of magnitude* in the intrinsic fitting time with respect to the original microstructure imaging techniques [9]. Moreover, AMICO guarantees convergence to a global minimum without any initialization procedure since its formulation is convex.

4.6 Conclusion

In this chapter, we have presented some background information on diffusion MRI, an MR technique that can measure the diffusion of water in biological tissue.

We have focused on applications of dMRI in the brain, where the anisotropy of diffusion in the white matter can be exploited 1) to map the neural connectivity in the brain through assessing the main orientation(s) of the fiber tracts voxelwise and 2) to characterize the microstructural properties of the white matter, like axon diameter and density. We have presented the state-of-the-art approaches to tackle both problems, providing the reader with an overview of their strengths and their main limitations.

Structured sparsity for spatially coherent fiber orientation estimation in diffusion MRI

5.1 Introduction

In this chapter, we propose a novel formulation to solve the problem of intra-voxel reconstruction of the fiber orientation distribution function (FOD) in each voxel of the white matter of the brain from diffusion MRI data.

A great variety of approaches have been proposed to tackle the problem of intra-voxel fiber orientation estimation (see chapter 4). Diffusion Tensor Imaging (DTI) [71] is one of the simplest and fastest reconstruction techniques since it only requires sampling 6 points of the q -space. However, it is by construction unable to model multiple fiber populations within a voxel and thus it is not valid in regions with crossings. Diffusion Spectrum Imaging (DSI) [84], on the other hand, is a model-free imaging technique known to provide good imaging quality. Yet, it requires strong magnetic field gradients and long acquisition times, needing typically 256 samples for a good reconstruction. As a consequence, it generally becomes too time-consuming to be of real interest in a clinical perspective. Accelerated acquisitions, relying on as few sampling points as possible while still sensitive to fiber crossings represent thus a major goal in the field.

In the last years, *spherical deconvolution* (SD) methods [5, 72, 73] have become very popular in the framework of local reconstruction since they can recover the fiber configuration with a relatively small number of points, typically from 30 up to 60. They consider that both anisotropy and magnitude of water diffusion in white matter (WM) are constant in the whole volume. Under this assumption, SD methods acknowledge the fact that the diffusion signal can be expressed as the convolution of a response function, or kernel, with the fiber orientation distribution function (FOD). The FOD is a real-valued function on the unit sphere that indicates the orientation and the volume fraction of the fiber populations in a voxel. The Constrained Spherical Deconvolution approach [5, 73] represents the first attempt to solve the ill-posed SD problem. It applies Tikhonov regularization, introducing a constraint on the ℓ_2 norm of the FOD, specially to ensure its positivity. Apart from the aforementioned work, most of the state-of-the-art methods to solve SD problems promote sparse regularization based on ℓ_1 minimization [6, 85, 86], where the

ℓ_1 norm is defined, for any real vector, as the sum of the absolute value of its coefficients. Yet, in [87] the authors acknowledge that ℓ_1 minimization is formally inconsistent with the fact that the volume fraction sum up to unity, and demonstrate the superiority of ℓ_0 -norm minimization. All these local reconstruction methods solve the FOD recovery problem for each voxel independently and thus, do not exploit the spatial coherence of the fiber tracts in the brain. A number of approaches have addressed this shortcoming by formulating the problem globally (simultaneously for all voxels) to be able to exploit the correlation between the different volumes. Some of them decouple the problem and propose a global denoising of the diffusion data prior to reconstruction [88, 89]. Another group of methods present a joint scheme for reconstruction and spatial regularization on the diffusion images at each q -space point. For instance, [90] propose a variational formulation to jointly estimate and regularize DTI to account for the effect of Rician noise in low SNR regimes, while the standard state-of-the-art minimization of the *total variation* (TV) semi-norm [91] of the diffusion images is used to denoise in [86, 92].

In this chapter, we propose a formulation that solves the fiber configuration of all voxels of interest simultaneously and imposes spatial regularization directly on the fiber space. This reconstruction allows us to exploit information from the neighboring voxels that cannot be taken into account by the existing state-of-the-art methods that approach fiber reconstruction independently in each voxel. The natural smoothness of the anatomical fiber tracts through the brain can be translated in a certain *spatial coherence* of the FOD in neighboring voxels. Accordingly, in the aim of recovering the global FOD field in all voxels, the present work leverages a reweighted ℓ_1 -minimization scheme to promote a spatially structured sparsity prior imposing spatial coherence. While the spatial regularization schemes proposed in [86, 90, 92] enforce sparsity of the images at each q -space point, our spatial regularization relates to the fundamental coherence between fiber directions – the FOD – in neighbor voxels, thus adding anatomically driven constraints. Our code is available at <https://github.com/basp-group/co-dmri> and it is distributed open-source.

The rest of the chapter is organized as follows: In section 5.2 we recall the framework for local FOD reconstruction through spherical deconvolution. We firstly introduce the local ℓ_0 algorithm in [87] and secondly propose a nonlocal method which solves for the FOD in all voxels simultaneously introducing spatial coherence of the fiber bundles orientation in neighboring voxels. We report and discuss results on both synthetic and real data in section 5.3. Conclusions and further work to be considered are examined in section 5.4.

5.2 Materials and Methods

5.2.1 dMRI framework for recovery of FOD via spherical deconvolution

In the SD framework, the intra-voxel structure estimation can be expressed through the FOD recovery problem in terms of the following linear formulation:

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\eta}, \quad (5.1)$$

where $\mathbf{x} \in \mathbb{R}_+^n$ stands for the FOD, $\mathbf{y} \in \mathbb{R}_+^m$ is the vector of measurements, Φ is the linear measurement operator and $\boldsymbol{\eta}$ is the acquisition noise. The reader can refer to [85] for a more detailed overview on SD methods and the formal equations describing the relationship between the FOD and the diffusion signal. We consider a dictionary Φ that spans a set of the Diffusion Basis Functions introduced in [6]. Each of these basis functions is generated by applying a different rotation to a kernel, which corresponds to the diffusion signal response to a single fiber. The set of available orientations represents a discretization of half of the unit sphere (\mathbb{S}^2), assuming antipodal symmetry in the diffusion signal. The diffusion signal can then be expressed as a linear combination of these basis functions, also referred to as the *atoms* of our dictionary Φ .

Prior constraints are essential to regularize a deconvolution problem like (5.1) in order to find a unique solution from an originally ill-posed problem. In the framework of the recently developed theory of *compressed sensing* (CS) [18, 20] sparsity priors are commonly used as regularizers to recover a signal from a set of undersampled measurements (see chapter 2). In formulation (5.1) the sparsity can directly be inferred from the small number of fiber directions of interest, in correspondence with the FOD coefficients. In this work, the method proposed in [87] is taken as the state-of-the-art algorithm in the framework of SD local methods for FOD recovery. For the sake of completeness, it is described in detail hereafter.

In [87], the authors propose to resort explicitly to the nonconvex ℓ_0 prior to solve for the FOD rather than to its convex ℓ_1 relaxation. A convex optimization problem for FOD reconstruction can be defined through a constrained formulation between adequate sparsity prior and data, also making use of a reweighted sparse deconvolution. The proposed minimization problem reads as:

$$\min_{\mathbf{x} \geq 0} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq k. \quad (5.2)$$

In (5.2), $\|\cdot\|_0$ represents the ℓ_0 norm (number of nonzero coefficients) and k acts as a bound on the expected number of fiber populations in a voxel. Since the ℓ_0 norm is nonconvex, a reweighted ℓ_1 -minimization scheme [24] is used in order to approach ℓ_0 minimization by

a sequence of convex weighted- ℓ_1 problems of the form:

$$\min_{\mathbf{x} \geq 0} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_{w,1} \leq k. \quad (5.3)$$

In (5.3), the ℓ_0 norm has been substituted by a weighted- ℓ_1 norm defined as $\|\mathbf{x}\|_{w,1} = \sum_i \mathbf{w}_i |\mathbf{x}_i|$. The algorithm alternates between estimating the solution at iteration t , $\mathbf{x}^{(t)}$, and redefining the weights essentially as the inverse of the values of the solution at the previous iteration $\mathbf{w}_i^{(t+1)} \approx 1/\mathbf{x}_i^{(t)}$ (see section 2.4.2). The use of these weights allows the algorithm to iteratively better estimate the nonzero locations and induces that, at convergence, the weighted- ℓ_1 norm mimics the ℓ_0 norm. Hence, formulation (5.2) promotes sparsity through a sequence of problems (5.3). Hereafter we will refer to this voxel-by-voxel method based on ℓ_2 and ℓ_0 priors as L2L0.

In the next subsection we describe an algorithm, inspired by L2L0, that exploits the anatomical coherence of the fiber tracts of the brain by promoting a *structured sparsity* prior on the FOD field. We show evidence that taking into account neighboring information through an appropriate prior directly on the object of interest improves significantly the results in comparison with solving for all voxels independently or using indirect spatial regularization schemes.

5.2.2 Spatial regularization through structured sparsity

In the aim of exploiting the spatial coherence of the fibers in the brain when recovering the local fiber configuration, we formulate a problem to solve the ensemble FOD field for all voxels simultaneously. To emphasize the fact that the minimization problem (5.2) is formulated separately for each voxel of the brain, we can rewrite it using the following notation:

$$\min_{\mathbf{x}^{(v)} \geq 0} \quad \|\Phi \mathbf{x}^{(v)} - \mathbf{y}^{(v)}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}^{(v)}\|_0 \leq k, \quad (5.4)$$

where $\mathbf{x}^{(v)} \in \mathbb{R}_+^n$ represents the real-valued FOD in the particular voxel indexed v . By concatenating all vectors $\mathbf{x}^{(v)}$ columnwise, one can build a matrix $\mathbf{X} \in \mathbb{R}_+^{n \times N}$, whose columns correspond to the FOD in each particular voxel. The elements of matrix \mathbf{X} will be indexed as \mathbf{X}_{dv} , each row d being associated with the atom of the dictionary oriented in direction indexed d , each column v being associated with voxel indexed v , $\mathbf{X}_{\cdot v} = \mathbf{x}^{(v)}$, as represented in Figure 5.3. N denotes the total number of voxels we want to recover the fiber configuration from. The rows of $\Phi \mathbf{X}$ represent the modeled diffusion images at each q -space point.

In our proposed formulation, a global data term is minimized adding a sparsity constraint that simultaneously promotes spatial coherence of the solution. Inspired by formulation (5.3), we adopt a procedure that consists in solving a sequence of problems of the

form:

$$\min_{\mathbf{X} \in \mathbb{R}_+^{n \times N}} \|\Phi \mathbf{X} - \mathbf{Y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{X}\|_{\mathbf{W},1} \leq K, \quad (5.5)$$

where the matrix $\mathbf{Y} \in \mathbb{R}^{m \times N}$ is formed by the concatenation of all N measurement column vectors: $\mathbf{Y}_{:,v} = \mathbf{y}^{(v)} \in \mathbb{R}^m$. The sensing matrix Φ is exactly the same as in (5.4) and $\|\cdot\|_{\mathbf{W},1}$ stands for a weighted ℓ_1 norm of a matrix defined as:

$$\|\mathbf{X}\|_{\mathbf{W},1} = \sum_{d,v} \mathbf{W}_{dv} |\mathbf{X}_{dv}|. \quad (5.6)$$

The following paragraphs are devoted to describe in detail the reweighting scheme and define the weighting matrix \mathbf{W} .

In a reweighted- ℓ_1 scheme, large weights will progressively tend to *discourage* nonzero entries whereas small weights will *promote* nonzero entries in the solution. The weighting matrix \mathbf{W} has the same dimension as \mathbf{X} and each of its entries acts as a weight for the corresponding entry of \mathbf{X} . The weights should still represent the inverse value of the associated entry at the previous iteration, so as to lead to an ℓ_0 -norm prior at convergence. However, a strong spatial coherence prior can actually be promoted by adapting the computation of the weights as follows. Our definition of the weights is driven by the underlying anatomical assumption that fiber bundles in neighboring voxels should have very close orientations as the trajectories are smooth (schematically represented in Figure 5.1). In terms of the FOD, this premise implies that neighbor voxels should bear *similar* directions.

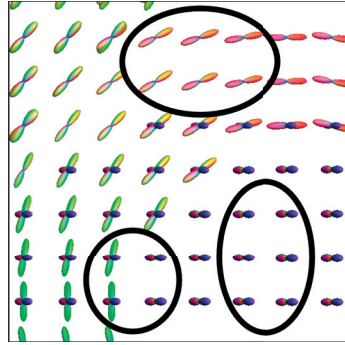


Figure 5.1: Synthetic FOD field in a representative 2D slice, which consists of two crossing fiber bundles. Due to the natural smoothness of the bundles, FODs in neighboring voxels are expected to contain similar peaks, as highlighted in the figure. *Figure published in [7].*

To translate this idea into a mathematical formulation of the weights we start by formally defining the concept of *neighborhood*. Since each atom of the dictionary represents a direction d on the half sphere, we define an *angular neighborhood* $\mathcal{N}(d)$ for each of them composed by the *closest atoms* (in terms of angular distance). In our implementation we have considered a maximal angular distance of 15° to delimit the neighborhood of each atom. Analogously, for each voxel v of the brain we define its *spatial neighborhood* $\mathcal{N}(v)$ as the group of 26 voxels that share either a face, an edge or a vertex with the voxel of interest v , commonly referred to as the 26-adjacent neighborhood [93]. A visual representation of

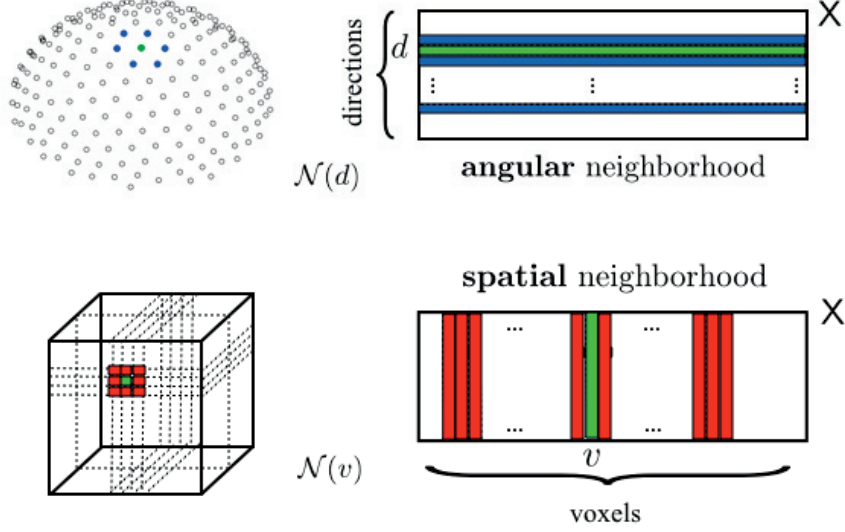


Figure 5.2: Top row: Schematic representation of a spatial neighborhood. On the left: Set of voxels representing the 3D-volume (brain) we want to solve for. Voxels in red configure the neighborhood $\mathcal{N}(v)$ for a particular voxel v , in green. On the right: Mapping of $\mathcal{N}(v)$ as a set of columns of matrix \mathbf{X} . **Bottom row: Schematic representation of an angular neighborhood.** On the left: Set of black circles representing the discretization of the half sphere chosen to build dictionary Φ . Points highlighted in blue configure the neighborhood $\mathcal{N}(d)$ for a particular direction d , in green. On the right: Mapping of $\mathcal{N}(d)$ as a set of rows of matrix \mathbf{X} . *Figure published in [7].*

both $\mathcal{N}(d)$ and $\mathcal{N}(v)$ is shown in Figure 5.2. For convenience, we define $\overline{\mathcal{N}(d)} = d \cup \mathcal{N}(d)$ and $\overline{\mathcal{N}(v)} = v \cup \mathcal{N}(v)$, the neighborhoods that include the central element. We then define the neighborhood of an element \mathbf{X}_{dv} as the entries of \mathbf{X} at the intersection of rows d and all its neighbor directions, and columns v and all its neighbor voxels: $\overline{\mathcal{N}(dv)} = \{(d', v'); d' \in \overline{\mathcal{N}(d)}, v' \in \overline{\mathcal{N}(v)}\}$, as it is schematically represented in Figure 5.3.

At each iteration, every element of the weighting matrix \mathbf{W}_{dv} is set as the inverse of an *average* of the absolute values that \mathbf{X} takes in the neighborhood of \mathbf{X}_{dv} in the previous iteration:

$$\mathbf{W}_{dv}^{(t+1)} = \left[\tau^{(t)} + \frac{1}{|\overline{\mathcal{N}(dv)}|} \sum_{d'v' \in \overline{\mathcal{N}(dv)}} |\mathbf{X}_{d'v'}^{(t)}| \right]^{-1}. \quad (5.7)$$

Consequently, at each iteration t , the weighting matrix $\mathbf{W}^{(t)}$ represents a *blurred* version of the current estimation of the solution $\mathbf{X}^{(t)*}$. In (5.7), we average over voxels, but sum over directions as all values in neighbor directions are interpreted as contributing to a single true local direction, in particular because the true direction does in general not coincide exactly to one of the discrete points of the sphere identifying our orientation dictionary.

*The values of the final solution are influenced by their weights, however they are not directly identified with them.

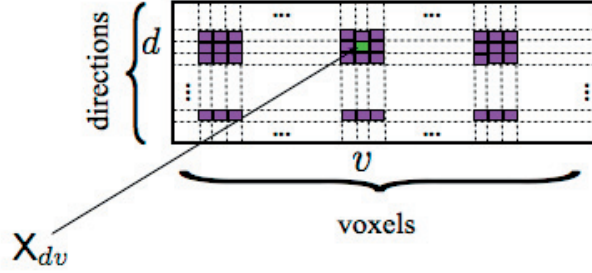


Figure 5.3: Schematic representation of the neighbourhood of element X_{dv} (in green), i.e. the elements of \mathbf{X} involved in the computation of weight W_{dv} . It includes coefficients corresponding to directions d and all its neighbours $\mathcal{N}(d)$ for voxel v and all its neighbours $\mathcal{N}(dv) = \{(d', v'); d' \in \mathcal{N}(d), v' \in \mathcal{N}(v)\}$. Figure published in [7].

This helps to stabilize the regularization and prevent the appearance of spurious peaks: fiber contributions are usually spread over a small angular support while spurious peaks are associated with isolated directions. To avoid infinite values for null averages, we add a stability parameter τ in the definition of the weights. We apply an homotopy strategy [94] and use a decreasing sequence $\{\tau^{(t)}\}$ in such a way that $\tau^{(t)} \rightarrow 0$ when $t \rightarrow \infty$. In the absence of any spatial constraint, $W^{(0)}$ corresponds to the matrix of all 1s and thus, the weighted ℓ_1 norm is the standard $\ell_{1,1}$ norm of a matrix, $\|\mathbf{X}\|_{W,1} = \|\mathbf{X}\|_{1,1}$.

The specific computation of the weights described in the former paragraphs encourages that neighbor voxels present the same or very close (neighbor) directions, imposing *structured sparsity* of the solution. Indeed, all entries corresponding to the neighborhood of an element contribute to its weight. Therefore those orientations that are “supported” by the surrounding voxels are reinforced, since they will be given a small weight compared to isolated directions that are not coherent with their environment. At convergence, our definitions (5.6) and (5.7) thus implement a spatially coherent version of the matrix ℓ_0 norm, i.e. the sum of the ℓ_0 norms of its columns. This reweighting scheme promotes a regularization that takes into account the true anatomy of the brain accounting for the fact that fiber populations present a coherent trajectory across voxels close to each other in the brain volume. This prior constitutes a powerful constraint that cannot be exploited when solving the problem independently for each voxel, like in (5.4).

The main steps of the *reweighting scheme* are reported in algorithm 3; in the remaining of the manuscript we will refer to it as $L2LO_{NW}$, in reference to the described *neighbor weighted* scheme. The reweighting process stops when the relative variation between successive solutions $\|\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}\|_2 / \|\mathbf{X}^{(t-1)}\|_2$ is smaller than some bound or after the maximum number of iterations allowed is reached.

Algorithm 3 Reweighted ℓ_1 minimization for global reconstruction of the FOD

Require: $\mathbf{Y} \in \mathbb{R}^{m \times N}$; $\Phi \in \mathbb{R}^{n \times m}$; K ; ν ; τ_{thr} ; N_{max} ; $\mathcal{N}(d)$, $d = 1, \dots, n$; $\mathcal{N}(v)$, $v = 1, \dots, N$
Ensure: FOD $\mathbf{X} \in \mathbb{R}_+^{n \times N}$

Initialize $t \leftarrow 0$; $\mathbf{X}^{(0)} = 0$; $\mathbf{W}^{(0)} \leftarrow 1$
while $\rho > \nu$ and $t < N_{\text{max}}$ **do**
 Solve:
 $\mathbf{X}^{(t)} \leftarrow \min_{\mathbf{X} \in \mathbb{R}_+^{n \times N}} \|\Phi \mathbf{X} - \mathbf{Y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{X}\|_{\mathbf{W},1} \leq K$
 Update $\mathbf{W}^{(t+1)}$
 Update $\rho = \|\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}\|_2 / \|\mathbf{X}^{(t-1)}\|_2$
 $t \leftarrow t + 1$
end while
 $\mathbf{X} \leftarrow \mathbf{X}^{(t-1)}$

5.2.3 Implementation details

To generate the dictionary Φ in our experiments, we estimated two different Gaussian kernels that model the diffusion signal in the regions of the brain corresponding to (i) white matter (WM) and (ii) partial volume with grey matter or cerebrospinal fluid (CSF). Modeling each kernel actually corresponds to estimating the three eigenvalues of the diffusion tensor. Grey matter and CSF are typically isotropic media. Consequently, their representative kernel is spherical – a tensor with three equal eigenvalues – and not sensitive to rotations. On the other hand, the kernel corresponding to the WM is anisotropic. Its response function was first estimated by fitting a tensor from the diffusion signal in those voxels with the highest fractional anisotropy (as expected to contain only one fiber population) and subsequently it was rotated in 200 different directions equally distributed on the sphere. Therefore, the final number of atoms of the dictionary used for this reconstruction is 201: 200 atoms corresponding to WM plus 1 isotropic atom modeling partial volume with CSF and grey matter.

Each weighted- ℓ_1 problem of the form (5.5) is solved using a forward-backward algorithm [95] in the context of proximal splitting theory (see chapter 2). To set a meaningful bound K we have followed the criterion that at convergence the weighted- ℓ_1 norm of a matrix, as defined in section 5.2.2, mimics the ℓ_0 norm – as in formulation (5.3) –. K is then heuristically fixed as $K = 3N$, as it represents a conservative bound on the total number of fiber orientations to be identified, computed as the number of voxels N times an average bound on the number of fiber orientation per voxel. We initialize $\tau^{(0)}$ as the variance of the solution after the first iteration $\mathbf{X}^{(0)}$ and, in subsequent iterations, we update $\tau^{(t+1)} = \beta \tau^{(t)}$ with $\beta = 10^{-1}$. Ideally $\tau^{(t)}$ should decrease to 0 but we heuristically fix a lower bound $\tau_{\text{thr}} = 10^{-7}$, above which significant signal components could be identified. Experiments show that for a convergence bound $\nu = 10^{-3}$ the reweighting process stops after a relatively small number of iterations, typically 4 or 5. In our simulations, ν is set to 10^{-3} and N_{max} to 10.

To extract the final fiber directions from the solution to algorithm 3 in every voxel

we perform a search for local maxima among all directions within a cone of 15° around every direction. In this entire process, we disregard the directions with contributions (i.e. coefficients) smaller than 10% of the maxima in order to filter out spurious peaks.

5.2.4 Phantom data

We perform our experiments using the phantom data used for the *HARDI reconstruction Challenge 2012* [96]. The public results in [96] allow us to compare the performance of L2L0_{NW} with other methods using different spatial regularization schemes – such as TV regularization mentioned above – with no need for their explicit implementation. The dataset is a $16 \times 16 \times 5$ volume that comprises 5 different fiber bundles that result in voxels with bending, crossing and kissing tracts. The response function of each bundle has been generated with a fractional anisotropy between 0.75 and 0.90 and the diffusion properties are constant along all its trajectory. More details on its geometry can be found in [96].

The signal is contaminated with *Rician noise* [97] as follows:

$$S_{\text{noisy}} = \sqrt{(S + \xi_1)^2 + (\xi_2)^2}, \quad (5.8)$$

with $\xi_1, \xi_2 \sim \mathcal{N}(0, \sigma^2)$ and $\sigma = S_0/\text{SNR}$ corresponding to a given signal-to-noise ratio (SNR) on the S_0 image. The quality of the reconstructions has been evaluated as a function of three different noise levels, i.e. $\text{SNR} = 10, 20, 30$ and 5 different q -space acquisition schemes (30, 20, 15, 10 and 6 samples), evenly spaced on half of the unit sphere.

5.2.5 Real Data

One HARDI[†] dataset was acquired at $b = 3000 \text{ s/mm}^2$ using 256 directions uniformly distributed on half of the unit sphere (as described in [98]), $\text{TR/TE} = 7000/108 \text{ ms}$ and spatial resolution $= 2.5 \times 2.5 \times 2.5 \text{ mm}$. To assess the robustness of L2L0_{NW} to different undersampling rates, the dataset has been retrospectively undersampled and three additional datasets have been created, consisting of only 30, 20 and 10 diffusion directions selected in order to be evenly spaced on half of the unit sphere using the tool *subsetpoints* which is available in the *camino* toolbox[‡]. We will refer to these four data sets as **hardi**₂₅₆, **hardi**₃₀, **hardi**₂₀ and **hardi**₁₀, respectively. The actual SNR in the $b = 0$ images, computed as the ratio of the mean value in a region-of-interest placed in the WM and the standard deviation of the noise estimated in the background, was about 30.

To evaluate the reconstructions from the undersampled real datasets, the metrics described in subsection 5.2.6 are computed considering the fully-sampled **hardi**₂₅₆ as the

[†]High-Angular Resolution Diffusion Imaging

[‡]www.camino.org.uk

golden truth, as it is suggested in [99].

5.2.6 Evaluation criteria

To evaluate the quality of the reconstructions we have focused on the performance of each method in both correctly assessing the number of fiber populations in each voxel and the angular accuracy in their orientation. In this work we adopted a set of metrics that was used to evaluate and compare all methods participating in the *HARDI reconstruction Challenge 2012* [96]. For consistency we have kept their notation to design the different quality indices. The *success rate* (SR_{\angle}) corresponds to the proportion of voxels in which a reconstruction algorithm correctly estimates the number of fiber populations. A fiber is considered to be correctly identified when an estimated fiber falls within a tolerance cone around a true fiber. To compare our results with different algorithms evaluated in [96], in this work the tolerance was set to 20° . *False positive* and *negative* rates (n_{\angle}^+ and n_{\angle}^- , respectively) are an average over all voxels of the number of over-/underestimated fiber populations per voxel.

The *angular accuracy* is measured through the *mean angular error* $\bar{\theta}$ (in degrees) averaged over all true fiber directions, where the angular error associated with each true fiber is formally defined as:

$$\theta = \frac{180}{\pi} \arccos(|\mathbf{d}_{\text{true}} \cdot \mathbf{d}_{\text{estimated}}|), \quad (5.9)$$

where \mathbf{d}_{true} and $\mathbf{d}_{\text{estimated}}$ are unitary vectors in the true fiber direction and the closest estimated direction. Note that indices SR_{\angle} , n_{\angle}^+ and n_{\angle}^- represent mean values over all voxels of interest, whereas $\bar{\theta}$ is computed voxelwise and we study its statistical distribution to evaluate the general angular accuracy of each reconstruction.

5.2.7 Experimental setup

In the next section, we evaluate the quality of reconstructions using L2L0_{NW}, both for numerical simulations and tests on real data. As shown in [87], L2L0 outperforms other state-of-the-art local methods that recover the FOD in the framework of spherical deconvolution. Consequently, we have chosen it as a benchmark to compare L2L0_{NW} with respect to methods that perform voxel-by-voxel reconstruction of the fiber configuration. We had access to the original implementation to run L2L0 reconstructions.

We also compare the performance of L2L0_{NW}, which jointly estimates the FOD and applies spatial regularization, with respect to applying first a nonlocal denoising procedure and subsequently perform local reconstruction. We have chosen an adaptation of the Linear Minimum Mean Squared Error (LMMSE) filter proposed in [88] to simultaneously filter all different gradient images. We use a publicly available implementation of the Joint

Anisotropic LMMSE filter[§] and subsequently apply L2L0 to reconstruct the FOD. We refer to this alternative as JAMMLSE+L2L0.

In addition, taking the advantage of the public results of the *HARDI reconstruction Challenge 2012* [96], we can compare the performance of L2L0_{NW} with a representative collection of state-of-the-art methods for simulations on phantom data. In particular, we are able to establish a comparison with other methods using different spatial regularization schemes – such as TV regularization mentioned above – with no need for an explicit implementation of these methods.

Our optimization code[¶] was implemented in MATLAB and run on a standard 2.4 GHz Intel Xeon processor. The non-optimized version of the code is able to reconstruct a whole brain volume of $106 \times 106 \times 51$ voxels within approximately 4 hours.

5.3 Results and discussion

5.3.1 Phantom data

In this subsection we start comparing in detail the performance for L2L0_{NW} relative to L2L0 and JAMMLSE+L2L0 for the phantom data set described in subsection 5.2.4. The performance of the three methods as a function of the undersampling rate in q -space is reported in Figure 5.4. We consider 5 different acquisitions schemes (30, 20, 15, 10 and 6 samples) and present results for two different noise levels, at SNR = 30 and SNR = 20. The plots demonstrate that L2L0_{NW} outperforms L2L0 and JAMMLSE+L2L0 for all number of samples, in both noise conditions. L2L0_{NW} exhibits an accurate reconstruction ($SR_{\angle} \geq 85$ and $\text{mean}(\bar{\theta}) \leq 6.5^\circ$), robust to noise for different undersampling regimes, down to 15 samples. Denoising high-SNR data prior to reconstruction, as it is done in JAMMLSE+L2L0, seems not to improve the quality of the reconstructions. Indeed, at SNR = 30, 20 JAMMLSE+L2L0 exhibits slightly worse results than L2L0 (moderately lower SR_{\angle} and $\bar{\theta}$). With high quality data (SNR = 30 and from 30 to 15 samples), the differences between the three methods are fairly mild. The superiority of L2L0_{NW} compared to L2L0 and JAMMLSE+L2L0 appears clearer as we move to higher undersampling regimes and SNR = 20, specially in terms of the ability of identifying the correct number of fibers (higher SR_{\angle}). The overall improvement in terms of the success rate is even more evident when we go down to 10 samples, where L2L0 and JAMMLSE+L2L0 exhibit a severe drop of the performance with $SR_{\angle} = 52$ (L2L0) and $SR_{\angle} = 50$ (JAMMLSE+L2L0) at SNR = 30 and $SR_{\angle} = 36$ (L2L0) and $SR_{\angle} = 38$ (JAMMLSE+L2L0) at SNR = 20, while $SR_{\angle} = 81$ (SNR = 30) and $SR_{\angle} = 72$ (SNR = 20) are obtained with L2L0_{NW}. We notice a significant deterioration of the reconstructions with all methods when decreasing the number of samples down to 6.

A more detailed analysis in severe noise conditions (SNR = 10) is presented in Fig-

[§]http://www.nitrc.org/projects/jalmmse_dwi/

[¶]Code is available at <https://github.com/basp-group/co-dmri>.

ure 5.5. The plots show an important difference between the performance achieved by L2L0, that solves the problem voxelwise, and L2L0_{NW} and JAMMLSE+L2L0 that take into account the correlation between voxels and directions. At SNR=10, the denoising step in JAMMLSE+L2L0, specially indicated to correct the effect of the Rician noise at low SNR regimes [88], improves drastically the quality of the reconstructions. In particular, the overall $\bar{\theta}$ performances differ significantly between L2L0 and JAMMLSE+L2L0, with an average enhancement of up to 5° in the mean $\bar{\theta}$ in different undersampling regimes. While in terms of angular resolution both L2L0_{NW} and JAMMLSE+L2L0 exhibit similar performance, L2L0_{NW} shows a higher SR_{\angle} down to 10 samples. In this noise setting, we analyze in detail the ability of correctly assessing the number of fibers through the false positives and negatives rates. Results show the effectiveness of the spatial regularization applied both in JAMMLSE+L2L0 and L2L0_{NW}, specially in avoiding overestimated directions (extreme decrease of n_{\angle}^+) even if the number of missed fibers (n_{\angle}^-) is also significantly decreased.

Plots analogous to Figures 5.4 and 5.5 can be found in [96], where an exhaustive comparison of all methods participating in the *HARDI reconstruction Challenge 2012*^{||} is presented. The performance of these algorithms is evaluated on the same phantom used in our simulations by computing the same quality metrics described in the present paper (SR_{\angle} , $\bar{\theta}$, n^+ and n^-). Figure 5.6 shows a comparison of the performance of L2L0_{NW} run with 15 samples with the following eight representative methods participating in the Challenge^{**}: (i) DTI_{neigh}, classical DTI method enhanced using contextual information [100]; (ii) L2-L1-DL, method using dictionary learning in the framework of ℓ_2 - ℓ_1 reconstruction [20]; (iii - iv) L2-L1-TV and L2-L1-TGV, using the ℓ_2 - ℓ_1 problem formulation and including spatial regularization schemes based on total variation and total generalized variation, respectively [86]; (v - vi) L2-L2 and NN-L2, based on ℓ_2 norm priors [6, 25]; (vii) DOT, classical diffusion orientation transform [101]; (viii) DSI_{LR}, classical DSI enhanced using Lucy-Richardson deconvolution [102]. For a more detailed explanation of each reconstruction method, you can refer to [96]. Direct quantitative comparisons with all these standard state-of-the-art algorithms is not straightforward from the results, since every method was tested using different sampling schemes (different number of samples and distribution of points). Yet, L2L0_{NW} can be positioned in the overall picture. In Figure 5.6, participant methods are sorted by the number of samples used for the reconstruction, increasing from left to right. The actual number of samples is indicated on the plot for every method. In mild noise conditions (SNR = 30), L2L0_{NW} is able to correctly assess the number of fibers in 85% of voxels ($SR_{\angle} = 85$) using as few as 15 signal samples and this quality appears comparable to the best SR_{\angle} scores obtained in the Challenge with methods using many more points (from 30 up to 257) to recover the fiber configuration. The superiority of L2L0_{NW} appears to be even more significant when a more noisy setting is considered. At SNR = 10, L2L0_{NW} using only 15 samples, shows the same quality of

^{||}http://hardi.epfl.ch/static/events/2012_ISBI

^{**}For the sake of consistency, all methods are named following the same notation as in [96].

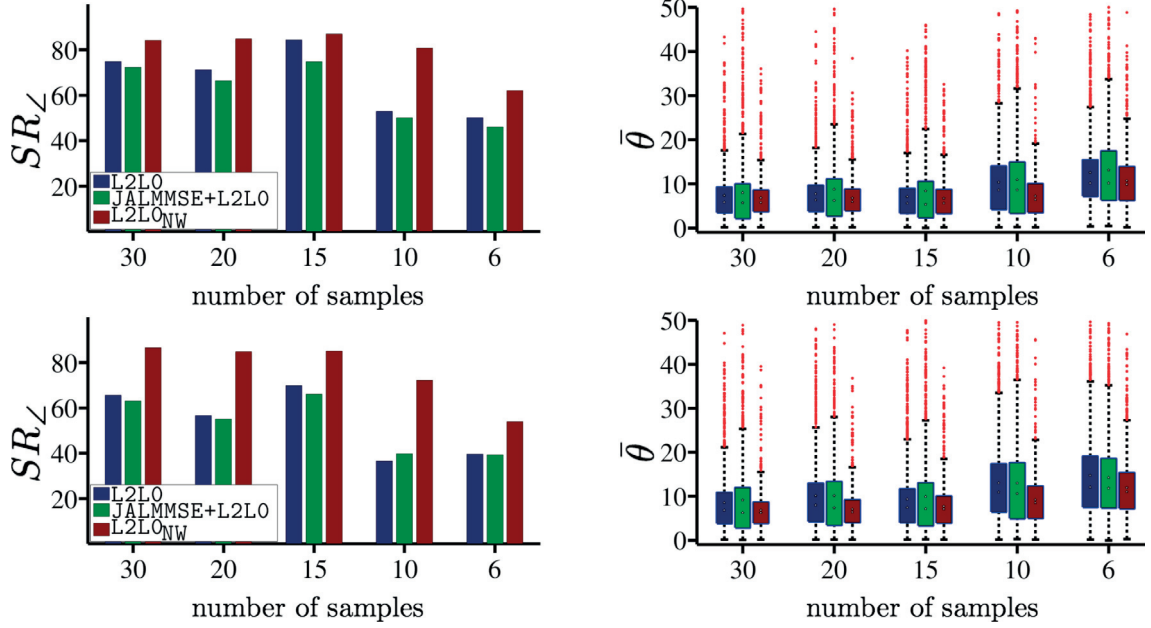


Figure 5.4: Comparison of SR_L and $\bar{\theta}$ between L2L0, JAMMLSE+L2L0 and L2L0_{NW} approaches. Experiments are performed on the phantom dataset used in [96] for a fixed SNR = 30 (top row) and SNR = 20 (bottom row). On the left, SR_L represents the *success rate*. On the right, the boxplot diagrams present the distribution of $\bar{\theta}$, with the edges of each box representing the 25th and 75th percentiles, the mean and median value appear as “square” and “circle” value and the outliers are plotted as red dots. *Figure published in [7].*

reconstruction, in terms of both SR_L and $\bar{\theta}$, as DSI using an exhaustive cartesian sampling scheme of 257 points. NN-L2 stands as the only method presenting slightly better results in terms of SR_L , yet, using 48 samples. Only with 15 samples L2L0_{NW} is able to attain comparable levels of performance, thus implying a speed-up factor of three. We pay special attention to the comparison with the rest of methods that promote any kind of spatial regularization. L2L0_{NW} with 15 samples ($SR_L = 85$ and $\text{mean}(\bar{\theta}) = 6.4^\circ$) outperforms L2-L1-TV, the method imposing TV regularization ([96]; see also [86]), in terms of success rate ($SR_L = 75$) and present similar average angular error ($\text{mean}(\bar{\theta}) = 6^\circ$), stressing the fact that the latter uses a sampling scheme with the double number of points (30 samples). Overall, we point out that all participant methods imposing spatial regularization (L2-L1-TV, L2-L1-TGV) use a significant amount of measurements (from 30 to 64 points) to recover the fiber configuration. The anatomical structured sparsity prior that we impose allows us to yield the same quality in the reconstructions using higher undersampling regimes.

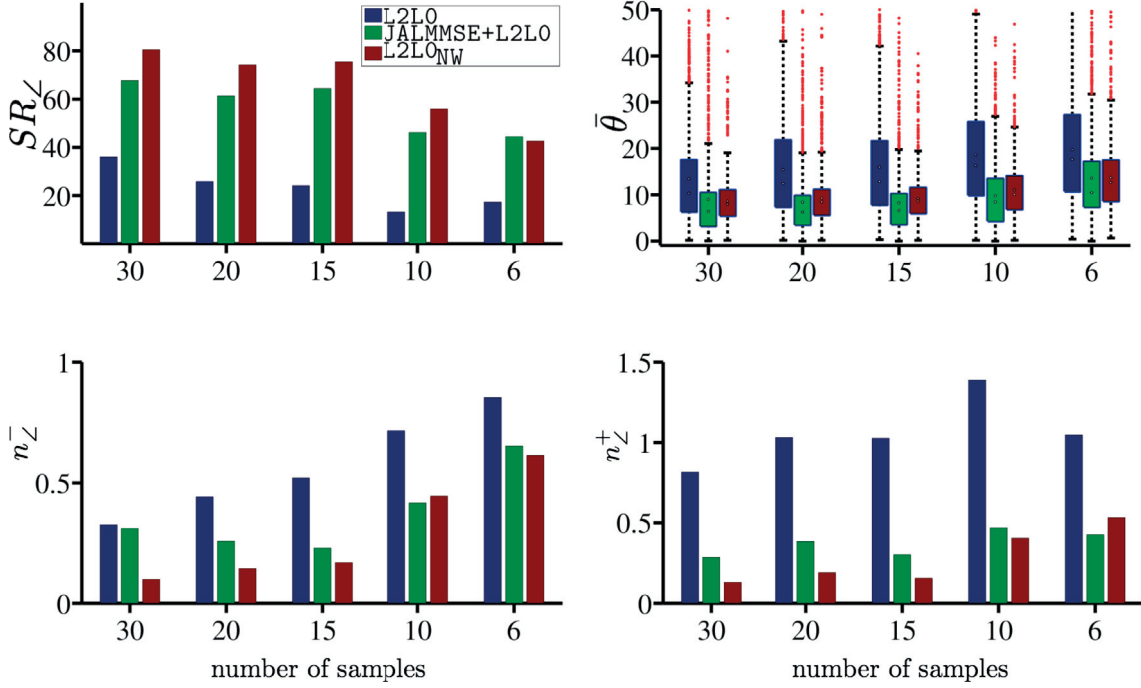


Figure 5.5: Comparison of SR_L , $\bar{\theta}$, n_L^- and n_L^+ between L2LO, JAMMLSE+L2LO and L2LO_{NW} approaches. Experiments are performed on the phantom dataset used in [96] for a fixed SNR = 10. On the top left, SR_L represents the *success rate*. On the top right, the boxplot diagrams present the distribution of $\bar{\theta}$ with the same conventions as for Figure 5.4. On the bottom row, n_L^- and n_L^+ represent the false negatives and positives rates. *Figure published in [7].*

5.3.2 Real Data

5.3.2.1 Quantitative comparison

In this subsection, we compare quantitatively the reconstructions obtained from undersampled real data (i.e. `hardi30`, `hardi20` and `hardi10`) to those with fully-sampled data (i.e. `hardi256`), considering the latter as ground-truth, for L2LO, JAMMLSE+L2LO and L2LO_{NW}. Results quoted next are in agreement with those obtained for numerical simulations on the phantom, confirming that L2LO_{NW} actually outperforms L2LO and JAMMLSE+L2LO. Bearing in mind that the actual SNR in the $b = 0$ images is about 30, results for JAMMLSE+L2LO and L2LO_{NW} appear in line with conclusions driven from the *HARDI Reconstruction Challenge 2012*, where it was shown that spatial regularization appeared to be effective also in low noise regimes, while merely denoising the images did not [96].

The average *mean angular error* ($\bar{\theta}$) using 30 samples was $13.9^\circ \pm 11.4^\circ$ (mean \pm standard deviation over WM voxels of the whole brain volume) for L2LO, $14.5^\circ \pm 10.8^\circ$ for JAMMLSE+L2LO and $7.8^\circ \pm 9.14^\circ$ for L2LO_{NW}. Reconstructions using 20 samples had an average error of $15.7^\circ \pm 11.2^\circ$ for L2LO, $16.7^\circ \pm 11.8^\circ$ for JAMMLSE+L2LO and $9.1^\circ \pm 9.6^\circ$ for L2LO_{NW}. When one goes down to 10 samples, reconstructions using L2LO and JAMMLSE+L2LO exhibit an angular error of $19.8^\circ \pm 11.25^\circ$ and $19.8^\circ \pm 12.0^\circ$, respectively,

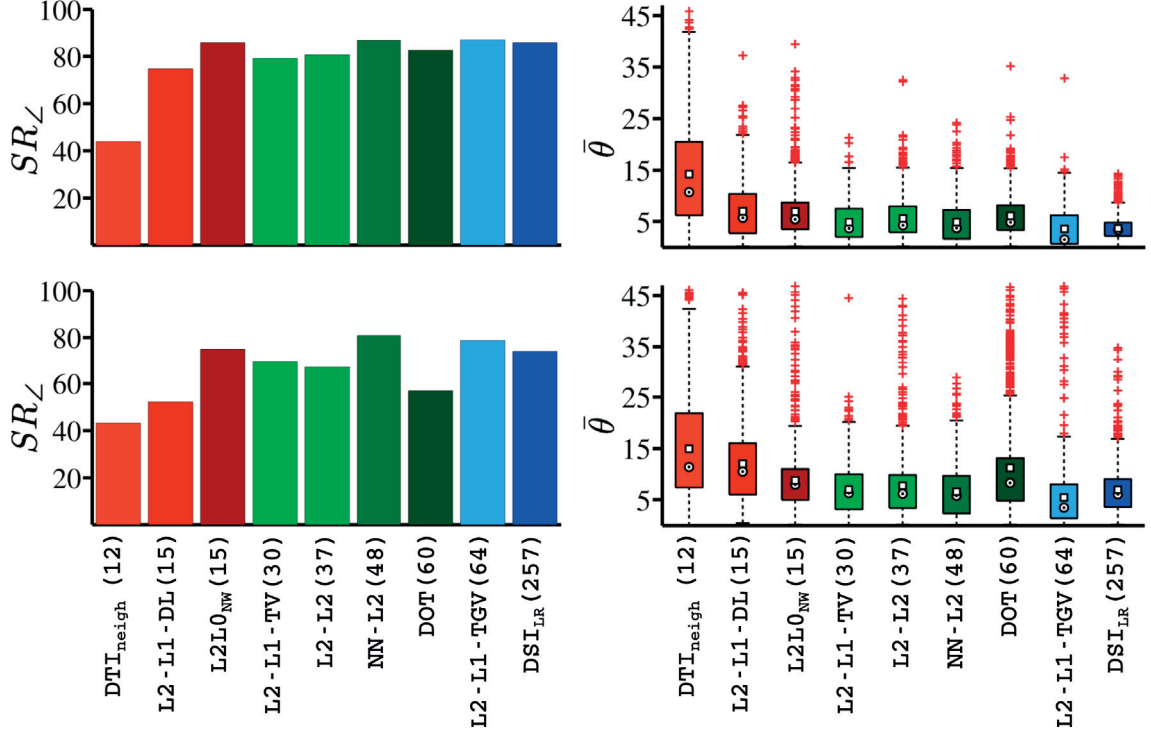


Figure 5.6: Comparison of SR_L and $\bar{\theta}$ between different reconstruction methods. Experiments are performed on the phantom dataset used in [96] for a fixed SNR = 30 (top row) and SNR = 10 (bottom row). On the left, SR_L represents the *success rate*. For the sake of comparison, the number of samples used for the reconstruction is reported in parentheses next to the name of each method. On the right, the boxplot diagrams present the distribution of $\bar{\theta}$, with the edges of each box representing the 25th and 75th percentiles, the mean and median value appear as “square” and “circle” value and the outliers are plotted as red dots. *Figure published in [7].*

which is already higher than the resolution of the spherical discretization defined by our dictionary; while the angular error for $L2L0_{NW}$ is $13.6^\circ \pm 10.5^\circ$. Results for the success rate are as well consistent with the results obtained in simulations. As in numerical simulations, the benefits of imposing a spatial regularization directly on the fiber orientations are more remarkable when we go to higher subsampling regimes. The SR_L was $31.1\% \pm 46.3\%$ for $L2L0$, $34.8\% \pm 47.6\%$ for $JAMMLSE+L2L0$ and $67.0\% \pm 47.0\%$ for $L2L0_{NW}$ with 30 samples; $27.9\% \pm 44.9\%$ for $L2L0$, $28.0\% \pm 45.0\%$ for $JAMMLSE+L2L0$ and $61.7\% \pm 48.6\%$ for $L2L0_{NW}$ at 20 samples. All methods present a degradation in the quality of their reconstructions when we go down to 10 samples, SR_L decreasing to $16\% \pm 36.6\%$ for $L2L0$, $18.8\% \pm 39.0\%$ for $JAMMLSE+L2L0$ and $40.6\% \pm 49.1\%$ for $L2L0_{NW}$.

Figure 5.7 illustrate the numerical results for one representative slice of the brain volume. The angular accuracy of each reconstruction is presented by plotting the mean angular error $\bar{\theta}$ per voxel in Figure 5.7. A map of the *number of false positives* and *false negatives* per voxel is used to illustrate the ability of each method of correctly assessing the number of fibers in Figure 5.8. The images show the superiority of $L2L0_{NW}$ with respect

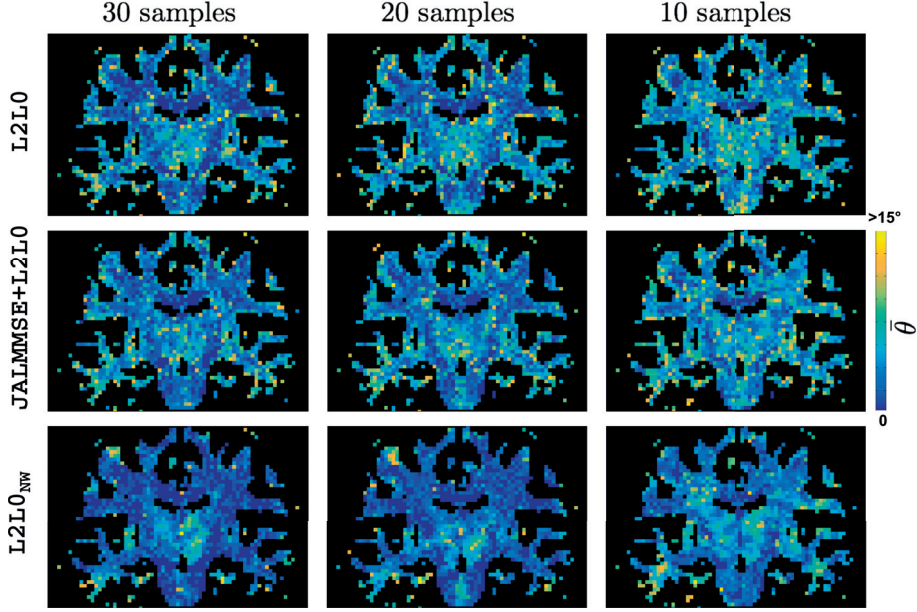


Figure 5.7: Angular accuracy (map of $\bar{\theta}$ per voxel) in real data between L2L0, JAMMLSE+L2L0 and L2L0_{NW} reconstructions with 30, 20 and 10 samples (`hardi30`, `hardi20`, `hardi10` datasets, respectively). *Figure published in [7].*

to L2L0 and JAMMLSE+L2L0, specially in those voxels close to the boundaries with the grey matter and the cerebrospinal fluid.

5.3.2.2 Qualitative comparison

The reconstructions^{††} of the FOD obtained with L2L0 and L2L0_{NW} for a significant slice of the brain in the corona radiata region are compared qualitatively in Figures 5.9 and 5.10. These plots show the robustness of each method to two different undersampling regimes, `hardi30` and `hardi10`. In the light of the quantitative results obtained for both phantom and real data and given the fact that qualitative differences between reconstructions using L2L0 and JAMMLSE+L2L0 are difficult to appreciate, we do not show qualitative results for JAMMLSE+L2L0. In all images, three meaningful regions with fiber bundle crossings have been highlighted. With 30 samples (Figure 5.9 corresponding to `hardi30`), the FODs reconstructed by L2L0_{NW} present neater and sharper profiles with less presence of spurious peaks than the ones reconstructed by L2L0. In addition, the fiber orientation distribution field reconstructed by L2L0_{NW} looks qualitatively smoother overall. As a consequence, fiber bundles are better defined through more clearly identified peaks. Plots in Figure 5.10 show reconstructions performed with only 4% of the original data (10 samples). In these images – corresponding to reconstructions with highly undersampled data – the

^{††}The images have been created using the tool *mrview* of *mrtrix*. This required the FOD from L2L0 and L2L0_{NW} to be previously converted to spherical harmonics.

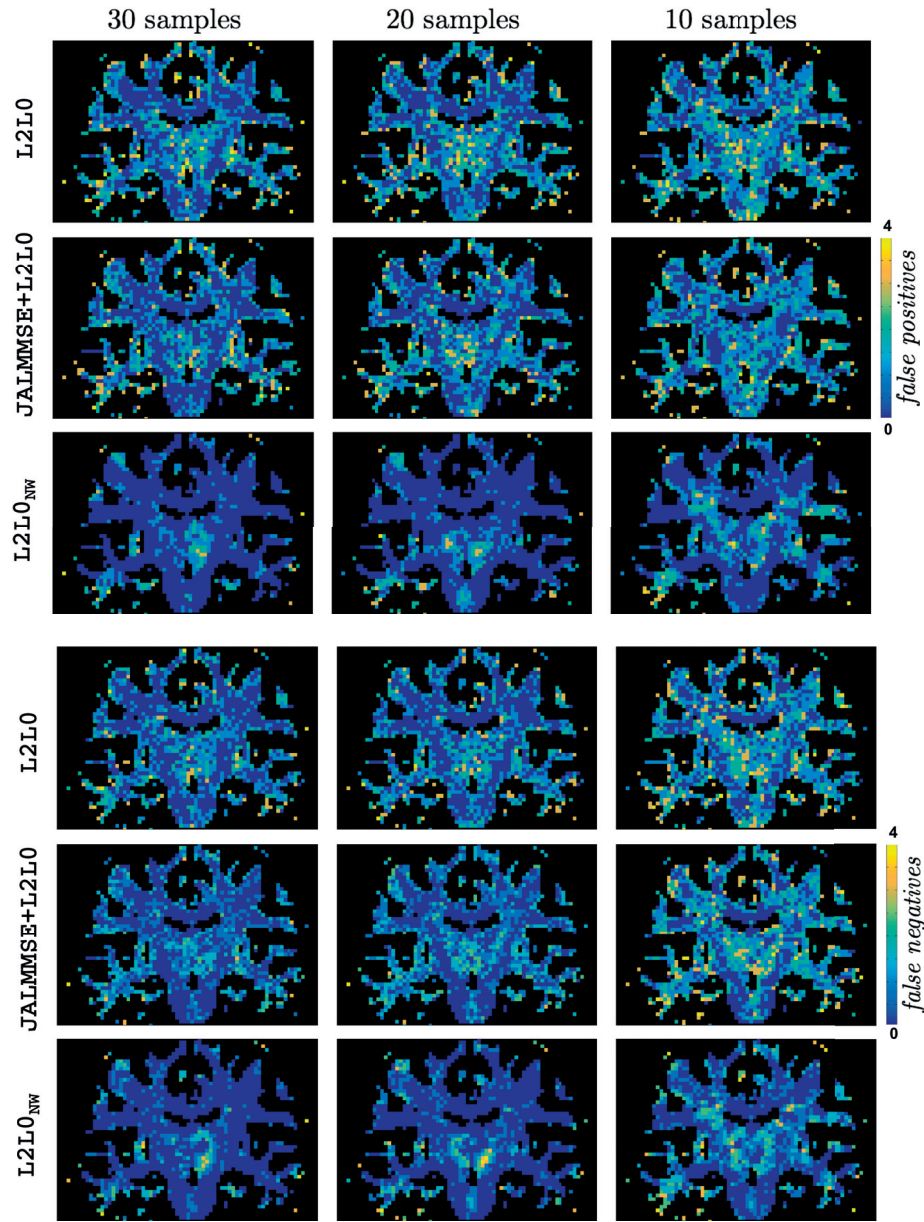


Figure 5.8: Ability of correctly assessing the number of fibers in real data between L2L0, JAMMSE+L2L0 and L2L0_{NW} reconstructions with 30, 20 and 10 samples (`hardi30`, `hardi20`, `hardi10` datasets, respectively). Map of number of *false positives* (top) and *false negatives* (bottom) per voxel. Figure published in [7].

above-mentioned qualitative differences between the two methods are confirmed and even more easily noticeable. As discussed in section 3.1, these differences can have a significant impact when applying tractography methods on these fiber orientation fields.

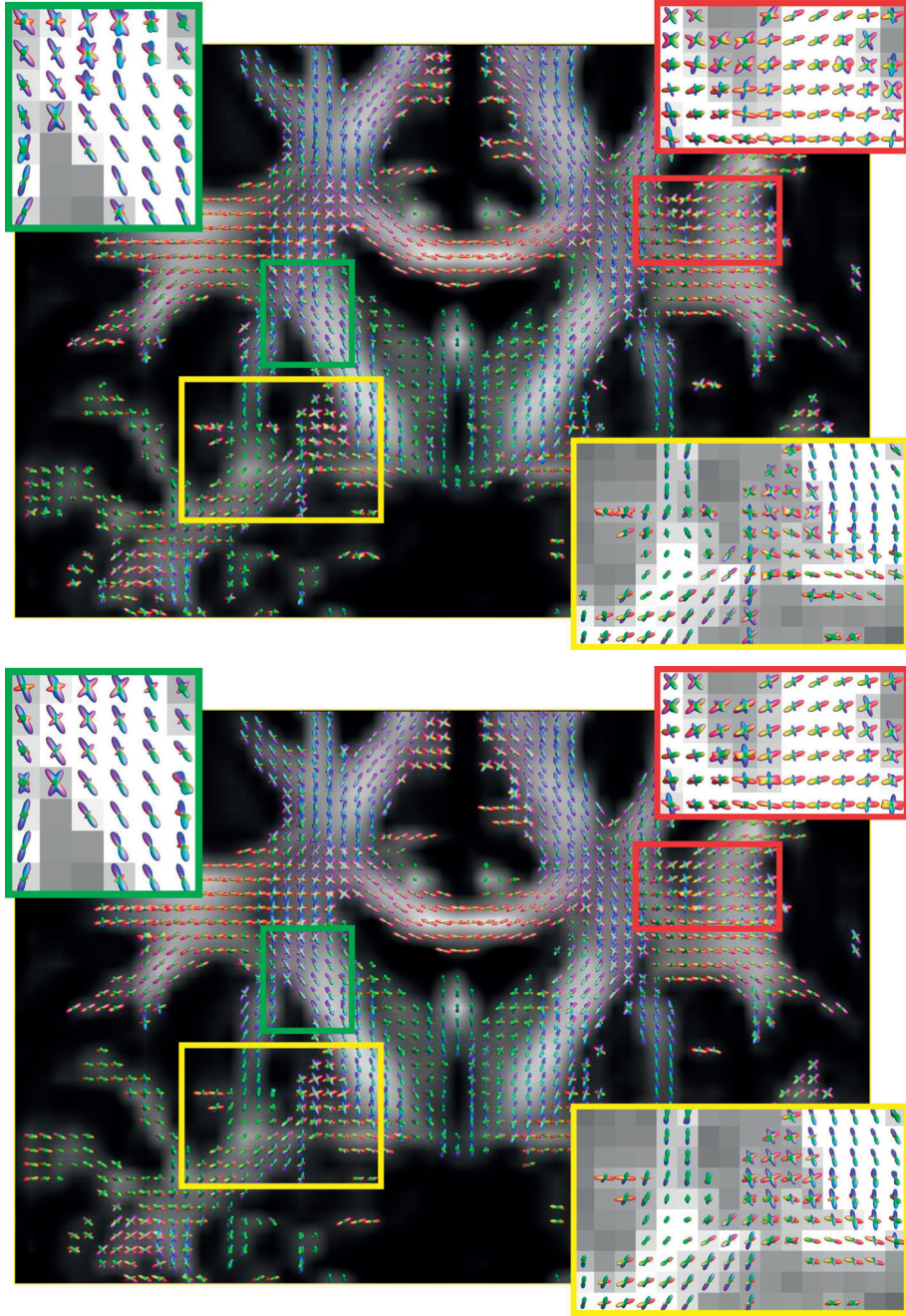


Figure 5.9: Qualitative comparison on HARDI human data. Reconstructions of the FODs in the corona radiata region are shown for L2L0 (top) and L2L0_{NW} (bottom) for **30 samples** superimposed to the FA map. *Figure published in [7].*

5.4 Conclusion

In this chapter we have proposed a novel algorithm to recover the intra-voxel FOD simultaneously for all voxels. The method leverages a spatially structured sparsity prior

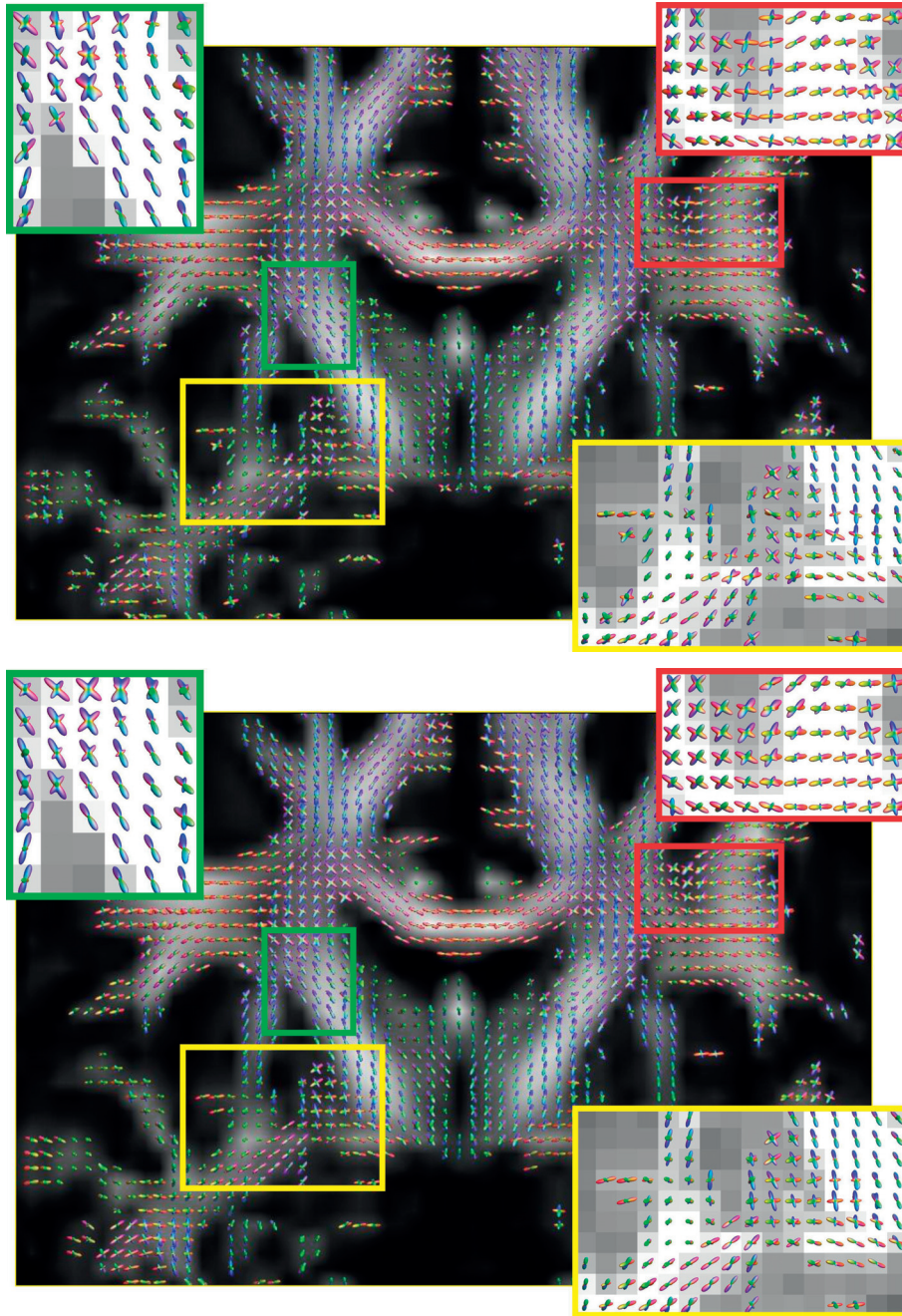


Figure 5.10: Qualitative comparison on HARDI human data. Reconstructions of the FODs in the corona radiata region are shown for L2L0 (top) and L2L0_{NW} (bottom) for **10 samples** superimposed to the FA map. *Figure published in [7].*

directly on the FOD, where the structure originates from the spatial coherence of the fiber orientation between neighbor voxels. We have made use of a reweighting scheme to enforce structured sparsity in the solution. We have shown through numerical simulations and tests on real data that this method outperforms a voxel-by-voxel reconstruction method

when assessing the correct number of fibers and the angular precision of their orientation. As shown in section 5.3, exploiting spatial information about the neighboring directions appears essential to ensure a stronger robustness to noise and ability to go to higher undersampling regimes, leading to accurate reconstructions with only 15 samples.

We also compare the performance of our proposed method with respect to applying first a nonlocal denoising procedure and subsequently perform local reconstruction. This comparison allows us to highlight the benefits of using a spatial regularization as in our approach as opposed to this decoupled strategy. As presented in simulations, our spatial prior on the FOD outperforms as well the empirical TV regularization of q -space images proposed by [86], being able to recover the fiber orientation distribution using fewer samples. Note that spatial regularization of the q -space images is actually complementary to our formulation and could be added as an additional prior to our method.

The regularization presented in this paper could as well be applied in a voxel-by-voxel configuration, redefining the weights in formulation (5.3) to account for the values of the FOD in a defined neighborhood. Preliminary investigations in this direction did not provide promising results. Fixing a single bound to estimate the number of fibers *separately* in every voxel of the brain appears to be too constraining. On the contrary, setting a bound on the total number of fibers of the whole volume and solving the problem for all voxels simultaneously leaves more freedom on the effective directions (number of nonzero coefficients) per voxel. Furthermore, future evolutions of this algorithm should enable undersampling in Fourier space (k -space) for each of the q -space images acquired. This combined $k - q$ -space sampling approach, along the lines of work by [86], will potentially enable a significant additional acceleration, in which context a voxel-by-voxel approach is not an option. Regarding computing resources, the memory requirements of a reweighting scheme to solve each voxel independently but using neighborhood information to define the weights would not differ from L2LO_{NW}, bearing in mind that the main operator Φ remains exactly the same for both formulations (5.3) and (5.5). In any case, the computation time of L2LO_{NW} is affordable for a single processor, as described in section 5.2.7.

In recent work [9], the authors present a general framework for Accelerated Microstructure Imaging via Convex Optimization (AMICO) to recover the microstructure configuration voxel-by-voxel in regions with one single fiber population. In the next chapter we consider the spatial coherence of the microstructural features of the fibers all over the brain with the aim of extending the AMICO framework to regions of the WM with multiple fiber populations and more complex configurations.

The work presented in this chapter has been published in [7, 8].

Fast microstructure estimation in regions with multiple fibers

6.1 Introduction

Most microstructure imaging techniques recover the microstructure properties by modeling the signal decay in different tissue compartments, e.g. axons, glial cells and extra-axonal space. These methods can infer not only the orientation of the main fiber population in a voxel, but also their microstructural properties, such as the apparent average diameter and density of the axons. For an exhaustive survey of the existing techniques in the field the reader can refer to chapter 4 and [75].

All the techniques mentioned in section 4.5 have demonstrated the practical possibility to estimate microstructural information from dMRI data and the estimated microstructural indices have been shown to agree very well with known anatomical patterns observed with histology [1, 82, 103]. However, they still suffer from sever limitations. On one hand, the nonlinear routines usually employed to fit these models are computationally very intensive and cause practical problems for their application in studies with several subjects. Secondly, they are only valid in regions with one single fiber population, making them inappropriate to characterize the microstructure of the majority of voxels in the brain. Furthermore, they still require acquisition times that make them difficult to implement in vivo in a clinical context.

Recently, Daducci et al. presented a flexible framework for Accelerated Microstructure Imaging via Convex Optimization (AMICO) [9] to reformulate these microstructure imaging techniques as linear systems that can be solved using convex optimization methods (see 4.5.1). The convex optimization framework enables to include prior information about the signal, such as positivity, as long as it is formulated as a convex constraint. Besides this flexibility, convex optimization methods are *fast* and many efficient numerical algorithms exist to solve them (see chapter 2). Despite the drastic improvement in speed, the current framework of AMICO replicates microstructure imaging techniques that so far are only valid in regions with one fiber population. Therefore, its use remains inadequate for many widespread regions of the brain with multiple fiber bundles. ActiveAx was recently extended to allow axon diameter mapping also in regions with crossing fibers [104] and

thus, overcome this limitation. Still, it requires about 1-hour scan and, thus, it is difficult to be routinely included in clinical studies.

In the first part of this chapter, we present a first preliminary extension of the AMICO framework to be able to recover microstructure parameters also in regions with multiple fiber populations using fast algorithms. Numerical simulations evidence the ability of our new approach to recover microstructure parameters in regions with crossing fibers.

In the second part of the chapter, we go one step further and propose a novel formulation that estimates the microstructure configuration and the fiber orientation simultaneously in all voxels as a global optimization problem, exploiting information from neighboring voxels that cannot be taken into account with existing techniques. Our preliminary results show the potential of our proposed method to enable robust reconstructions from a reduced number of diffusion measurements, thus leading to faster acquisitions, too.

6.2 AMICO_x

6.2.1 Materials and Methods

The reconstruction problem for microstructure features from diffusion data accounting for multiple fibers is presented here as an extension of AMICO [9]. In this preliminary work we focus on extending the formulation for the ActiveAx model [1] to enable axonal diameter mapping in case of multiple fiber populations within a voxel.

As in original AMICO, the reconstruction problem is decoupled into two simpler sub-problems. First, the number and orientation of the fiber populations $\boldsymbol{\mu}_i \in \mathbb{S}^2$ in each voxel is estimated. This can be achieved using any of several reconstruction methods, such as the standard Constrained Spherical Deconvolution (CSD) method [73]. Secondly, the linear operator Φ to express ActiveAx as a linear system is built from different sub-matrices:

$$\Phi = [\Phi_1^r | \Phi_1^h | \dots | \Phi_M^r | \Phi_M^h | \Phi^i]. \quad (6.1)$$

In equation (6.6), sub-matrices $\Phi_i^r \in \mathbb{R}^{m \times N_r}$ and $\Phi_i^h \in \mathbb{R}^{m \times N_h}$ model, respectively, the intra-axonal and extra-axonal contributions to the diffusion signal along the direction $\boldsymbol{\mu}_i (i = 1, \dots, M)$. M stands for the total number of detected fiber populations in the voxel. Each atom in sub-matrices Φ_i^r models the diffusion signal corresponding to water molecules *restricted* within parallel cylinders of a specific diameter. Alternatively, the atoms in sub-matrices Φ_i^h describe the *hindered* space between the axons. N_r and N_h represent, respectively, the number of different axon radii and hindered environments considered to build the dictionary. Sub-matrix $\Phi^i \in \mathbb{R}^m$ has a single atom that models the isotropic contribution corresponding to the CSF (note that it is orientation-independent). The signal response matching both restricted and hindered water diffusion in a voxel is estimated using the same models and parameter set as in [9]. For further details, the reader can refer to the original manuscript and find a specific description of these models

in [75].

The microstructure recovery problem is then solved as a Tikhonov-regularized least squares problem as follows:

$$\min_{\mathbf{x} \geq 0} \frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \frac{1}{2} \|\mathbf{x}\|_2^2, \quad (6.2)$$

where $\|\cdot\|_2$ is the standard ℓ_2 norm and the parameter $\lambda \geq 0$ controls the trade-off between data regularization terms. The microstructure indices of interest defined by Alexander et al. [1] can be estimated *for each individual fiber population* from the recovered coefficients \mathbf{x} by partitioning them as $[\mathbf{x}_1^r | \mathbf{x}_1^h | \dots | \mathbf{x}_M^r | \mathbf{x}_M^h | \mathbf{x}^i]$, corresponding to the contributions of hindered, restricted and isotropic compartments from every fiber bundle. In every voxel, the *intra-axonal volume fraction* ν' indicates the ratio between restricted and hindered compartments; and for each of the fiber populations i ($i = 1, \dots, M$), the *mean axon diameter index* a'_i is expressed as a weighted average of the coefficients corresponding to restricted water diffusion, \mathbf{x}_i^r :

$$\nu' = \frac{\sum_{i=1}^M \sum_{j=1}^{N_r} x_{i_j}^r}{\sum_{i=1}^M (\sum_{j=1}^{N_r} x_{i_j}^r + \sum_{j=1}^{N_h} x_{i_j}^h)} \quad (6.3)$$

$$a'_i = \frac{\sum_{j=1}^{N_r} 2R_j x_{i_j}^r}{\sum_{j=1}^{N_r} x_{i_j}^r}, \quad (6.4)$$

where $R_j, j \in \{1, \dots, N_r\}$ indicates the radius of the cylinder corresponding to the j -th atom in Φ_i^r . $x_{i_j}^r$ (alternatively, $x_{i_j}^h$) denotes the contribution corresponding to the j th restricted (alternatively, hindered) atom oriented along direction $\boldsymbol{\mu}_i$. We underline the fact that a'_i denotes a *weighted* mean axonal diameter and is not expected to correspond to the *true* mean axonal diameter over the axonal diameter distribution corresponding to the i th fiber bundle (see its precise definition in 4.5 or refer to [1]). However, by abuse of language, we often refer to a'_i as mean axonal diameter when discussing our results.

Hereafter, to make results easier to interpret for the reader, we refer to the original AMICO formulation [9] as AMICO₁ and to its extended version for multiple fibers as AMICO_X. In the next section, the performance of both formulations is compared through numerical simulations.

6.2.2 Experiments

To evaluate the effectiveness of AMICO_X, we tested it on synthetic data generated using the Monte-Carlo diffusion simulator system available in Camino [105], with the imaging protocol corresponding to a gradient strength $G_{max} = 140\text{mT/m}$ with 270 measurements divided into 3 shells with b -values = $\{1930, 3090, 13190\}\text{s/mm}^2$, corresponding to $G = \{140, 131, 140\}\text{mT/m}$, $\delta = \{10.2, 7.6, 17.7\}\text{ms}$, $\Delta = \{16.7, 45.9, 35.8\}\text{ms}$ and same

$TR/TE = 5000/60\text{ms}$ for all images. In all experiments when building the linear operators, we considered $N_r = 10$ different axon radii in a range of $0.1 - 8.5 \mu\text{m}$, and $N_h = 7$ different hindered environments corresponding to intra-axonal volume fractions from 0.3 to 0.9. No partial volume with CSF was included in the simulations, thus we did not consider any isotropic compartment in our dictionary ($\Phi^i = \emptyset$). Since the goal of this study remains to evaluate the ability of AMICO_x to correctly retrieve the microstructure indices of interest, we have chosen the regularization parameter λ that gives the minimal average relative error when fitting the mean axonal index over a set of 5 different pair of WM substrates. As a result, λ was fixed to 0.25.

We first simulated raw voxels with two fiber populations crossing at different angles (from 30° to 90°). Each fiber population consisted of a distribution of different axon diameter, as done in [1], and several WM substrates were tested. For each configuration, different relative ratios of the two populations were evaluated. In each case, the mean and standard deviation of the estimated microstructural parameters was computed over 1000 repetitions, contaminating the signal with independent Rician noise realizations corresponding to $\text{SNR} = 30$, and compared them to the ground-truth. The estimation of the fiber orientations was performed using standard CSD [73] and the CSD peak estimation using the toolbox MRTrix* with 90 measurements corresponding to the outer shell. For compactness, only results corresponding to relative volume fractions $f_{r_1} = \{0.5, 0.5\}$ and fiber population with 2 different radii – gamma distributions with parameters $(3.27, 4.9 \cdot 10^{-7})$ and $(4.82, 2.6 \cdot 10^{-7})$, respectively – corresponding to average axon diameters about 5.6 and 3.6 micrometers are reported here. Results on the other substrates are consistent.

6.2.3 Results and discussion

To show evidence of the need to consider more than one fiber population in the model, we fitted AMICO₁ in the experimental settings described above to assess the impact of using this single-fiber model in regions with more than one fiber population. In these experiments, the atoms of the dictionary were oriented in the direction estimated with DTI, as in the original formulation [9]. The estimated microstructure indices (mean axon diameter and intra-cellular volume fraction) are compared with the ground-truth in Figure 6.1. AMICO₁ assumes that the fibers inside the voxel follow only one direction. Results show that making such an assumption in voxels that actually contain more than one fiber leads to erroneous estimation of microstructural properties. Mean axonal diameter appears overestimated whereas the intra-axonal volume fraction is underestimated; and the absolute error increases with the crossing angle of the ground-truth fibers. As expected, in the AMICO framework, the importance of correctly estimating the number of fiber bundles in order to choose a correct model appears to be crucial.

Figure 6.2 compares the microstructure parameters estimated from AMICO_x with the

*<http://www.nitrc.org/projects/mrtrix/>

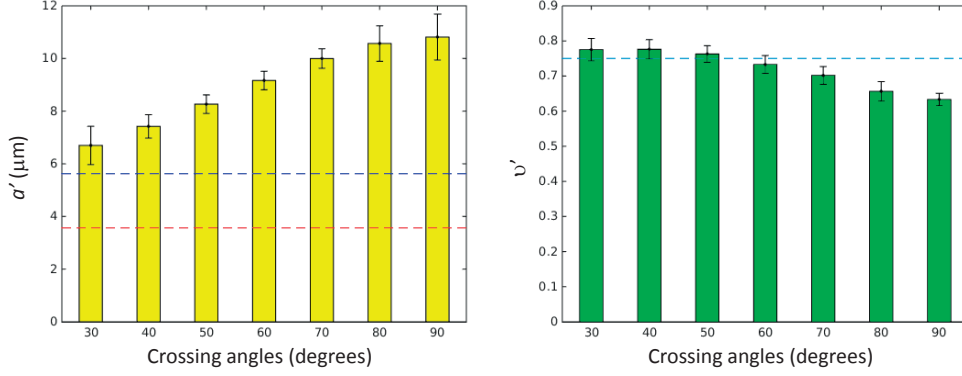


Figure 6.1: Performance of AMICO₁ on 2-fiber synthetic substrates as a function of the crossing angle between the fibers. Plots show the mean and standard deviation of the estimated *mean axonal diameter* (left) and *intra-axonal volume fraction* (right). Dashed lines represent the ground-truth values for the two populations.

ground-truth as a function of the crossing angles between the two fiber populations. The intra-cellular volume fraction can be estimated very accurately for all crossing angles (slightly over-estimated by about 4%). The mean axonal diameter of the two fiber populations can be as well estimated pretty robustly (both slightly under-estimated) for all crossing angles. However, when the two orientations are too close ($\approx 30^\circ$), the errors as well as the standard deviations of the estimates with respect to the ground-truth increase. These results are in line with (and slightly improve) those previously reported in [104]. The higher instability shown at 30° can be well related to the performance of CSD in the peak-detection step. While the average angular error committed over the 1000 repetitions in crossings from 90° to 40° is less than 2° , CSD often identifies spurious peaks as true fiber directions for angles crossing at 30° , leading to a more unstable behavior and higher average angular error.

Figure 6.3 illustrates the impact of the angular inaccuracy of the orientation of the fiber populations μ_i on the estimation of the microstructure indices. In a substrate with two fibers crossing at a fixed angle of 60° , one of the directions used to build the dictionary was deviated from 1° to 10° from the actual orientation of the fiber. The intra-cellular volume fraction can be estimated accurately for all angular deviations, up to 10° . The estimation of the mean axonal diameter degrades progressively, yet absolute errors are smaller than $1\mu\text{m}$ for angular deviations up to 7° . These results are in-line with the angular accuracy of AMICO₁ [9].

Lastly, the proposed model was tested also in a voxel with 3 non-coplanar fiber populations, as a proof of concept that evidences its generalization to multiple fiber crossings. In this experiment, the crossing angle between two of the fibers was fixed to 90° and the angle between the third one and the others varying between 30° and 90° . Only results corresponding to a crossing of two populations with an average axon diameters about 5.6 and one of about 3.6 micrometers are reported. Again, results with different substrates

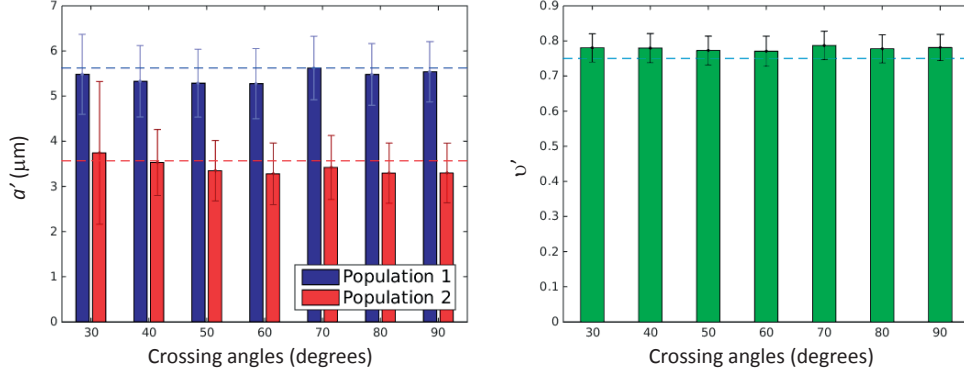


Figure 6.2: Performance of AMICO_x on 2-fiber synthetic substrates as a function of the crossing angle between the fibers. Plots show the mean and standard deviation of the estimated *mean axonal diameter* (left) and *intra-axonal volume fraction* (right) for the two different fiber populations. Dashed lines represent the ground-truth values for the two populations.

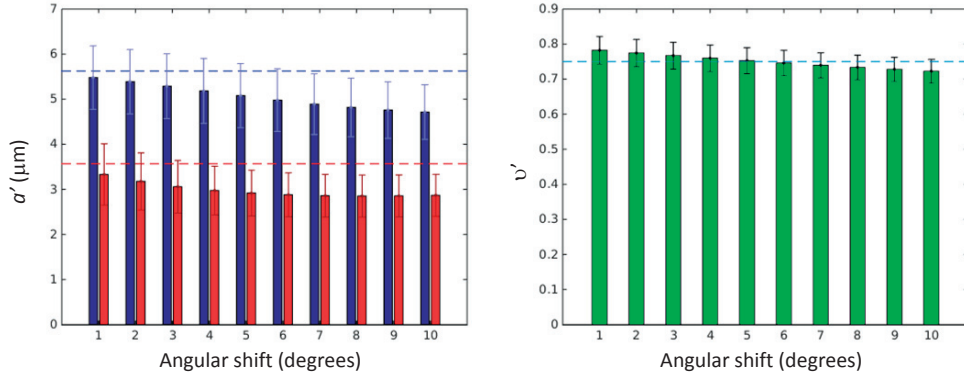


Figure 6.3: Robustness to inaccuracies in the estimation of μ_2 . Plots show the estimated *mean axonal diameter* (left) and *intra-axonal volume fraction* (right) as a function of the angular deviation of the estimated direction μ_2 with respect to the actual orientation of fiber population 2. Dashed lines represent the ground-truth values for the two populations.

are consistent. Figure 6.4 compares the estimated microstructure features with the ground truth as a function of the crossing angles between the 3 estimated populations.

The non-optimized version of the code, implemented in MATLAB and run on a standard 2.70GHz Intel Core i7-3740QM processor, is able to fit the model in approximately 3.7ms/voxel. Therefore, AMICO_x still enables a drastic reduction of the computation time to solve the microstructure imaging problem as well in regions with multiple fiber populations compared to other nonlinear routines, such as ActiveAx, which take $\approx 20\text{s/voxel}$ to fit its model [9].

So far, we have extended the original AMICO framework, that enables fast axonal diameter mapping with ActiveAx [1], to include crossing fiber populations within a voxel. Our results show through numerical simulations that AMICO_x is indeed able to robustly

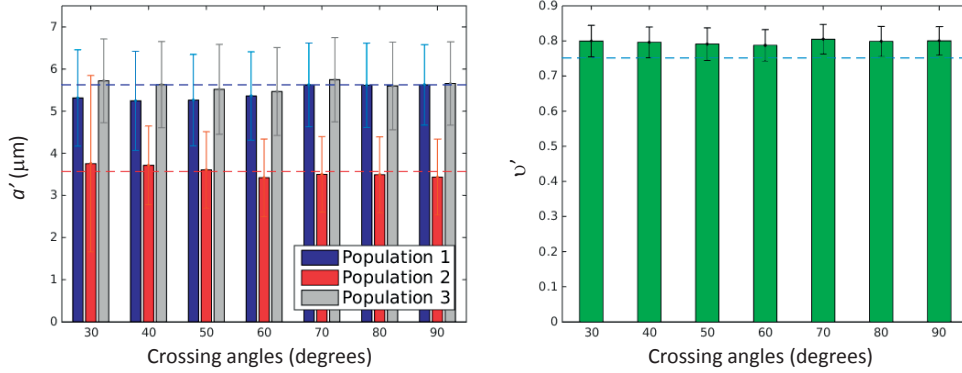


Figure 6.4: Performance of AMICO_x on 3-fiber synthetic substrates as a function of the crossing angle between the fiber population 1 and 2. The crossing angle between populations 1 and 3 is fixed to 90°. Plots show the mean and standard deviation of the estimated *mean axonal diameter* (left) and *intra-axonal volume fraction* (right) for the three different fiber populations. Dashed lines represent the ground-truth values for the three populations.

estimate the microstructure parameters, provided the number and orientation of the fiber populations in a voxel is correctly estimated (up to $\approx 7^\circ$ of angular accuracy). We have as well shown how, thanks to the fast convex optimization methods, AMICO_x enables a reduction of the computation time by orders of magnitude with respect to other microstructure imaging techniques also in voxels with complex fiber configurations. The extended data model of AMICO_x can be fitted fast and accurately in all voxels of the brain, as in [104], thanks to the generalization of the original formulation to environments with multiple fibers.

In the next section, a reformulation of the AMICO framework is proposed to enable microstructure reconstructions from a reduced number of measurements, thus leading to faster acquisitions.

6.3 AMICO_{SAM}

6.3.1 Materials and Methods

With the goal of enabling reconstructions in a highly-undersampled regime and thus speeding-up the acquisition time, we decide to impose a stronger regularization prior to exploit the smoothness of the fiber characteristics throughout the brain. To do so, we reformulate the microstructure recovery problem for the whole field of voxels simultaneously and therefore, we can take advantage of the neighbor information that cannot be taken into account when considering the problem independently in each voxel. Inspired by AMICO_x for ActiveAx to cope with multiple fiber populations and adapting the weighting scheme described in chapter 5 to exploit neighboring information, axon diameter mapping is presented here as a sequence of weighted linear problems as follows:

$$\min_{\mathbf{X} \in \mathbb{R}_+^{N_c \times N}} \|\Phi \mathbf{X} - \mathbf{Y}\|_2^2 + \lambda \|\mathbf{X}\|_{\mathbf{W},1}. \quad (6.5)$$

In (6.5), matrix $\mathbf{Y} \in \mathbb{R}^{m \times N}$ contains the m diffusion measurements from all N voxels of the image. The dictionary:

$$\Phi = [\Phi_1 | \dots | \Phi_n] = [\Phi_1^r | \Phi_1^h | \dots | \Phi_n^r | \Phi_n^h] \quad (6.6)$$

is a concatenation of sub-dictionaries Φ_i , built as in section 6.2.1, each modeling the intra- and extra-axonal contributions to the dMRI signal along direction μ_i , and $\{\mu_1, \dots, \mu_n\}$ is the set of discrete orientations uniformly distributed on the half-sphere used in the reconstruction.

Each recovered coefficient is associated to a restricted (or hindered) compartment with axonal diameter (or perpendicular diffusivity) m , oriented in direction d for voxel v , and therefore can be indexed using a triple index dvm . Throughout this section we refer to each coefficient as an element of a three dimensional tensor \mathcal{X} ; however, formally we organize them in matrix form – $\mathbf{X} \in \mathbb{R}_+^{N_c \times N}$, with $N_c = n \cdot (N_r + N_h)$ – so that we can easily express the linear convolution as a matrix product, see Figure 6.5. These coefficients allow us to recover, for every voxel:

- (i) The orientation of the fiber bundles.
- (ii) The microstructure indices defined in [1], which are estimated in every voxel and for each individual fiber population from the recovered coefficients \mathbf{X} , as described in formulations (6.3) and (6.4).

We highlight that, unlike the original AMICO framework, this new formulation enables the simultaneous estimation of the number of fiber populations present in the voxel, their orientation(s) and their microstructure characteristics.

In (6.5), we minimize a global data term and a sparsity constraint that simultaneously promotes spatial coherence of the solution, like in formulation (5.5). In this case, the sparsity stems from the small number of fiber directions of interest. Indeed, among the considered set of n discrete directions we expect only a few of them to have non-negligible values. $\|\cdot\|_{\mathbf{W},1}$ stands for the weighted ℓ_1 -norm which, by properly designing \mathbf{W} , induces spatially structured sparsity in the solution following the principle introduced in chapter 5 (see equation (5.6) and explanation below). In this case, the weights \mathbf{W} allow us to exploit neighborhood information and promote coherence in the so-called “*spatial-angular-microstructure*” (SAM) space. Since each restricted atom is associated to a specific diameter, we define its microstructural neighborhood $\mathcal{N}(m)$ as the two atoms with the closest bigger and smaller associated diameter. Figure 6.6 shows a visual representation of $\mathcal{N}(m)$, analogous to $\mathcal{N}(d)$ and $\mathcal{N}(v)$ defined in chapter 5. Similarly, for each hindered

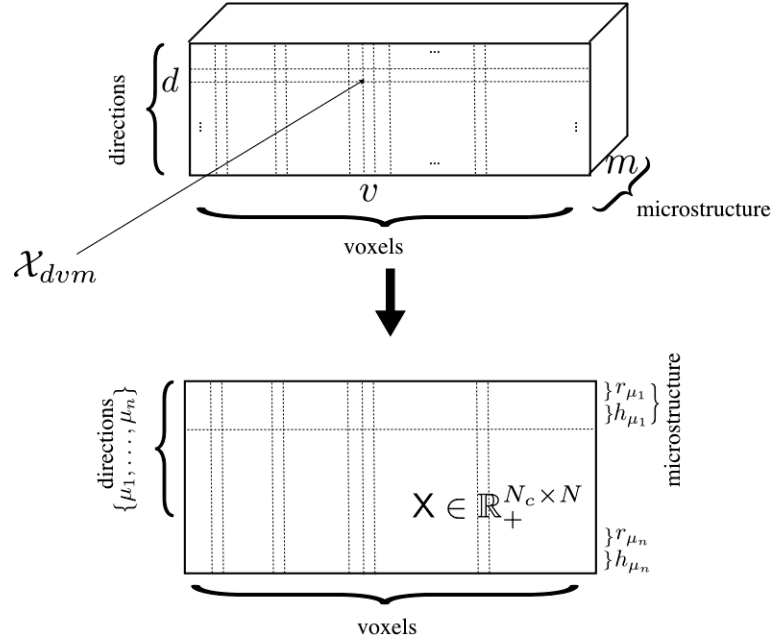


Figure 6.5: Schematic representation of the unknown in tensor and matrix form

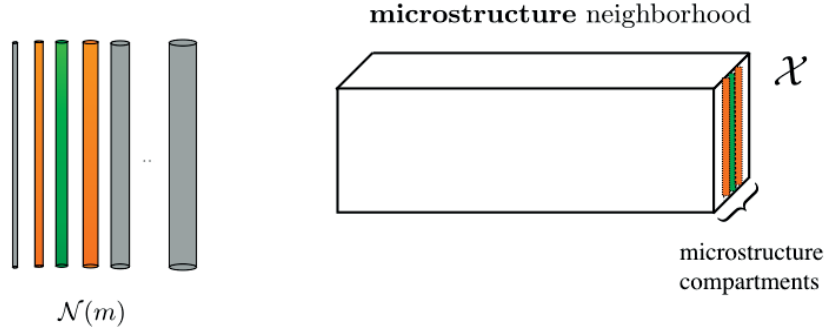


Figure 6.6: Schematic representation of the a neighborhood in the microstructure space. On the left: Set of cylinders representing different restricted compartments. Cylinders in orange configure the neighborhood $\mathcal{N}(m)$ for a particular compartment m , in green. On the right: Mapping of $\mathcal{N}(m)$ as a set of columns in tensor \mathcal{X} . Figure analogous to 5.2 to illustrate how we have added the microstructure dimension to our formulation in chapter 5.

atom associated to a specific perpendicular diffusivity, its neighborhood is defined as the two atoms with the closest larger and smaller diffusivity.

The weights regularize the coefficients in the microstructure dimension depending on their spatial position and orientation, and promote anatomical coherence among them according to the underlying fiber structure. Formally, they are defined adding the microstructural dimension to equation (5.7), as follows:

$$\mathcal{W}_{dvm}^{(t+1)} = \left[\tau^{(t)} + \frac{1}{|\mathcal{N}(v)|} \sum_{d'v'm' \in \mathcal{N}(dvm)} |\mathcal{X}_{d'v'm'}^{(t)}| \right]^{-1}. \quad (6.7)$$

In (6.7), the weights \mathcal{W} are expressed in tensor form to make use of the triple indexing. However, we formally organize them in matrix form \mathbf{W} so that $\|\mathbf{X}\|_{\mathbf{W},1}$ is well defined.

The whole reweighting scheme imposes a structured sparsity of the solution, encouraging that neighbor voxels present the same or close – neighbor – directions and microstructure. Indeed, these weights promote that in every voxel, the directions and the microstructure features are coherent with their environment.

The role and updating strategy for parameter τ in (6.7), as well as the main steps of the reweighting scheme are the same as described in chapter 5.

6.3.2 Experiments

To evaluate the effectiveness of $\text{AMICO}_{\text{SAM}}$, we tested it on synthetic data generated using the Monte-Carlo diffusion simulator system available in Camino [105], with the same imaging protocol detailed in section 6.2.2. In order to assess the robustness to data undersampling, we simulated different datasets using a decreasing number of measurements, from 270 (fully-sampled) to 54 (as typical HARDI protocols). The microstructural indices estimated with $\text{AMICO}_{\text{SAM}}$ were compared to the original voxelwise fitting, AMICO_{X} , and results are reported as averages over 10 Rician noise realizations (SNR=30).

Since the goal of this study is mainly to investigate the behavior of $\text{AMICO}_{\text{SAM}}$ in highly undersampled regimes, in all experiments, we tuned λ empirically to get the minimum average absolute error of the estimated parameters over several WM substrates.

We present results on two different kind of simulated substrates. We first show simulations considering only the signal decay due to water diffusion in the intra-axonal compartments. We simulated a field of 10×10 raw voxels corresponding to two fiber populations, consisting of cylinders with different radii distributions – as in [1] – crossing at given angles ($30^\circ - 90^\circ$). As in the previous section, we performed the following tests with different WM substrates but, for compactness, we report only results corresponding to populations with mean axonal diameter 2.9 and 5.2 μm . Results on other substrates were consistent. Secondly, to account for a more realistic situation, we simulated the signal corresponding to a field of 10×10 raw voxels with two fiber populations crossing, including this time the contribution due to water diffusing outside the axons.

In all simulations, when building the linear operators, we considered $n = 50$ directions uniformly distributed on the half sphere and $N_r = 7$ different axon radii in a range of 1 - 7 μm . In simulations where the extra-axonal compartment is not taken into account, we considered $N_h = 0$, otherwise $N_h = 7$ corresponding to volume fractions from 0.3 to 0.9.

6.3.3 Results and Discussion

Simulations without extra-axonal compartment

Figures 6.7 and 6.8 show the reconstructions for the two methods corresponding to 270 and 54 samples, respectively. The estimated microstructure indices (mean axonal diameter and population ratio) are shown as a function of the crossing angle between the fibers; dashed lines correspond to the ground truth values. In a fully-sampled context (Figure 6.7), there are no significant differences between the quality of the two reconstructions other than an increasing variability in the estimation with AMICO_X compared to AMICO_{SAM}, mainly when fibers get closer (30°). However, in a highly undersampled regime (Figure 6.8), AMICO_{SAM} continues to estimate accurately and precisely the population relative ratio whereas AMICO_X shows difficulties when the fibers are too close ($\leq 45^\circ$). Regarding the estimation of the mean axonal diameter with 54 samples, AMICO_X shows very unstable reconstructions and does not reliably assess the true values. On the other hand, AMICO_{SAM} is able to disentangle the two fiber populations and shows more accurate estimations at all crossing angles, even if we can observe a lower precision (higher std) in the estimates with respect to the reconstructions with a fully-sampled dataset.

Simulations including extra-axonal compartment

Figures 6.9 and 6.10 show the microstructure estimates for two WM substrates with mean axonal diameters 5.6 and 2.8 μm and an intra-axonal volume fraction of 0.75. These reconstructions appear in general more unstable than those considering only the intra-axonal compartment (Figures 6.7 and 6.8). We have observed that in some cases the algorithm “confuses” the restricted atoms associated to the highest diameters and hindered atoms, leading to an increasing general variability in the microstructure estimates.

In a fully-sampled context (Figure 6.9), there are no significant differences in the quality of the volume fraction and population relative ratio estimates between the two methods, even if we can observe a higher variability in the intra-axonal volume fraction estimated with AMICO_{SAM}. However, regarding the mean axonal diameters, AMICO_X estimates them accurately for both fiber populations, whereas AMICO_{SAM} consistently underestimates the highest mean axonal diameter by about 10%. We suspect the higher variability that AMICO_{SAM} shows in the estimation of the intra-axonal volume fraction compared to AMICO_X is linked to this underestimation of the mean axonal diameter, pointing to a confusion between some restricted/hindered atoms that we referred to in the former paragraph.

In a highly-undersampled regime (Figure 6.10), AMICO_X shows a very unstable reconstruction of the mean axonal diameters for both populations and an increasing degradation of the population relative ratio estimates as the two fibers get closer. On the other hand, the mean axonal diameters, relative population ratio and intra-axonal volume fraction recovered with AMICO_{SAM} are consistent for all crossing angles and coherent with the es-

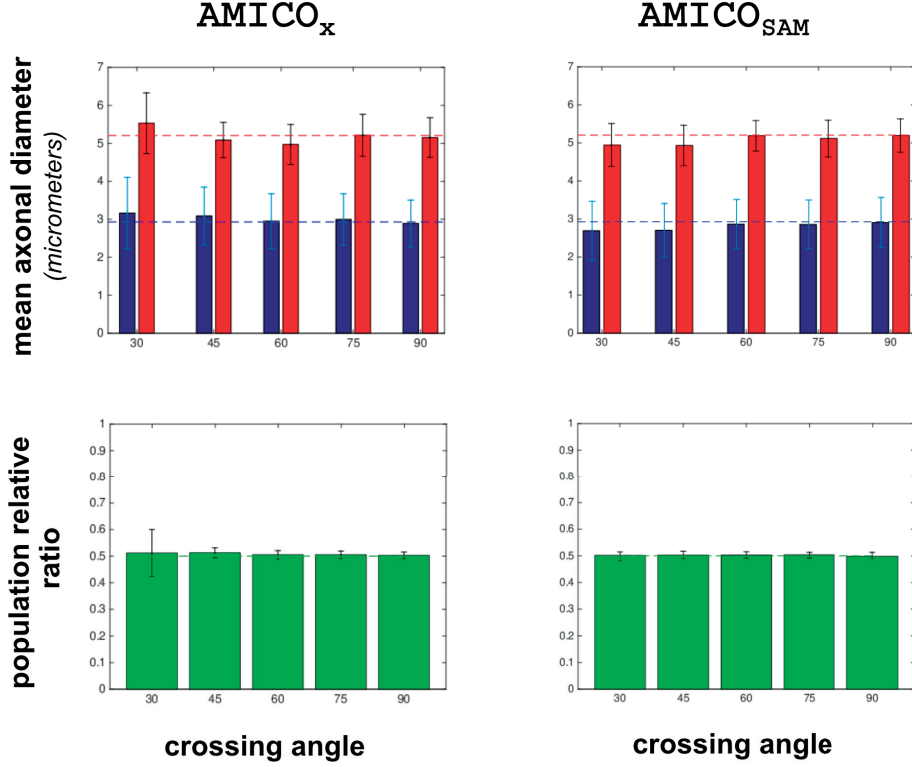


Figure 6.7: Detailed comparison of the mean axonal diameter index and population relative ratio estimated with AMICO_x (left) and AMICO_{SAM} (right), as a function of the crossing angle between the fiber populations. Results correspond to the fully-sampled protocol with **270 measurements**, considering only the water diffusion inside the axons.

timations using 270 samples, even if we observe an increased variability (higher standard deviation). Considering that AMICO_x is a 2-step procedure, a deeper analysis to better understand the origin of the instability of its highly-undersampled reconstructions should be carried out. Indeed, we should evaluate the angular errors committed when assessing the main fiber orientations and their impact on the estimation of the microstructure indices. Anyway, the inability of disentangling the 2 fiber populations at 75° suggests that the main source of error comes from the minimization problem itself, and not from the angular error.

Overall, we highlight that even if AMICO_{SAM} consistently underestimates the highest mean axonal diameter, it is able to correctly disentangle the two fiber populations for all crossing angles also in a highly-undersampled regime and that the three microstructure features are robustly estimated as we undersample.

6.3.4 Limitations and future work

In this section we have presented AMICO_{SAM}, that provides a flexible framework to solve both the fiber orientation and the microstructure recovery problems simultaneously

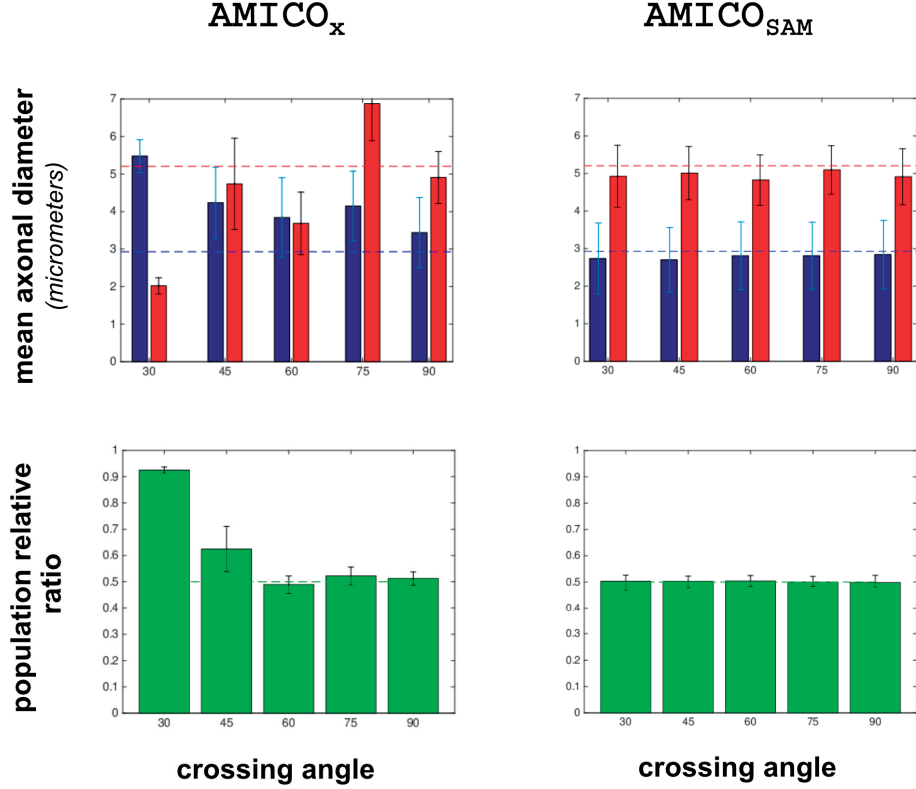


Figure 6.8: Detailed comparison of the mean axonal diameter index and population relative ratio estimated with AMICO_x (left) and AMICO_{SAM} (right), as a function of the crossing angle between the fiber populations. Results correspond to an undersampled protocol with **54 measurements** ($\approx 20\%$ of the fully-sampled protocol), considering only the water diffusion inside the axons.

while applying a regularization in three spaces: spatial, angular and microstructural. The preliminary work that we have presented so far shows robust reconstructions from a reduced number of diffusion measurements, opening the door to fast acquisitions for diffusion microstructure imaging. However it still shows many limitations that need to be overcome and appear as the target for our future work.

Study different sampling schemes/sequences

In this work, we have simulated data assuming PGSE sequences with the same sampling protocol used in [1], that corresponds to 3 shells with a maximum gradient strength $G_{max} = 140\text{mT/m}$. Different sequences (other than PGSE) and sampling protocols need to be investigated. Indeed, problem (6.5) is ill-posed since operator Φ is extremely ill-conditioned. One of the strategies to improve the conditioning of our linear operator Φ would be to decrease the correlation between its columns by designing a different sampling protocol. We are confident that this would help the algorithm in correctly distinguishing between atoms corresponding to intra- and extra-axonal compartments.

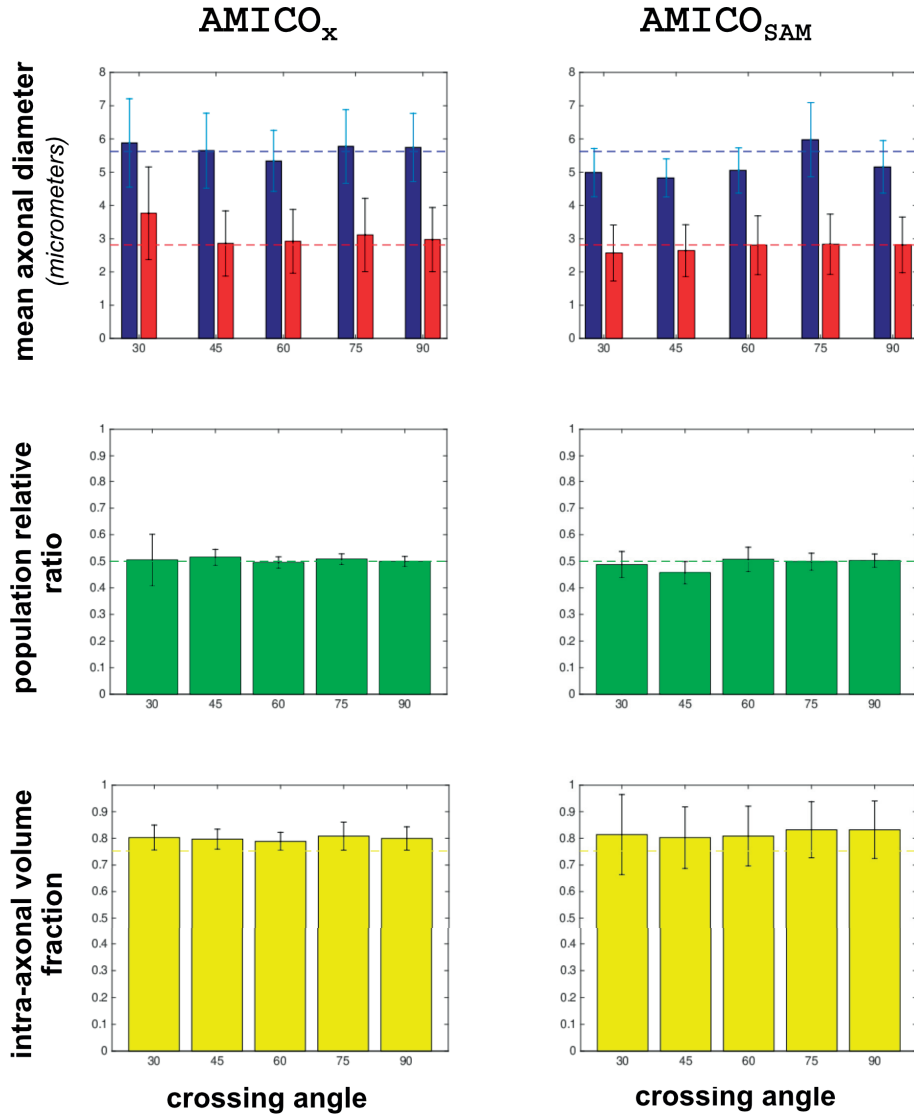


Figure 6.9: Detailed comparison of the mean axonal diameter index and population relative ratio estimated with $AMICO_x$ (left) and $AMICO_{SAM}$ (right), as a function of the crossing angle between the fiber populations. Results correspond to the fully-sampled protocol with **270 measurements**

Study the design of the dictionary

Intrinsic parameters of the algorithm, such as the chosen number of directions on the half-sphere n or the number and the range of the intra- and extra-axonal compartments have a direct impact on the complexity of the dictionary. In close relation with the previous point, the design of the dictionary should be optimized so that the linear inverse problem to be solved is the least ill-posed as possible.

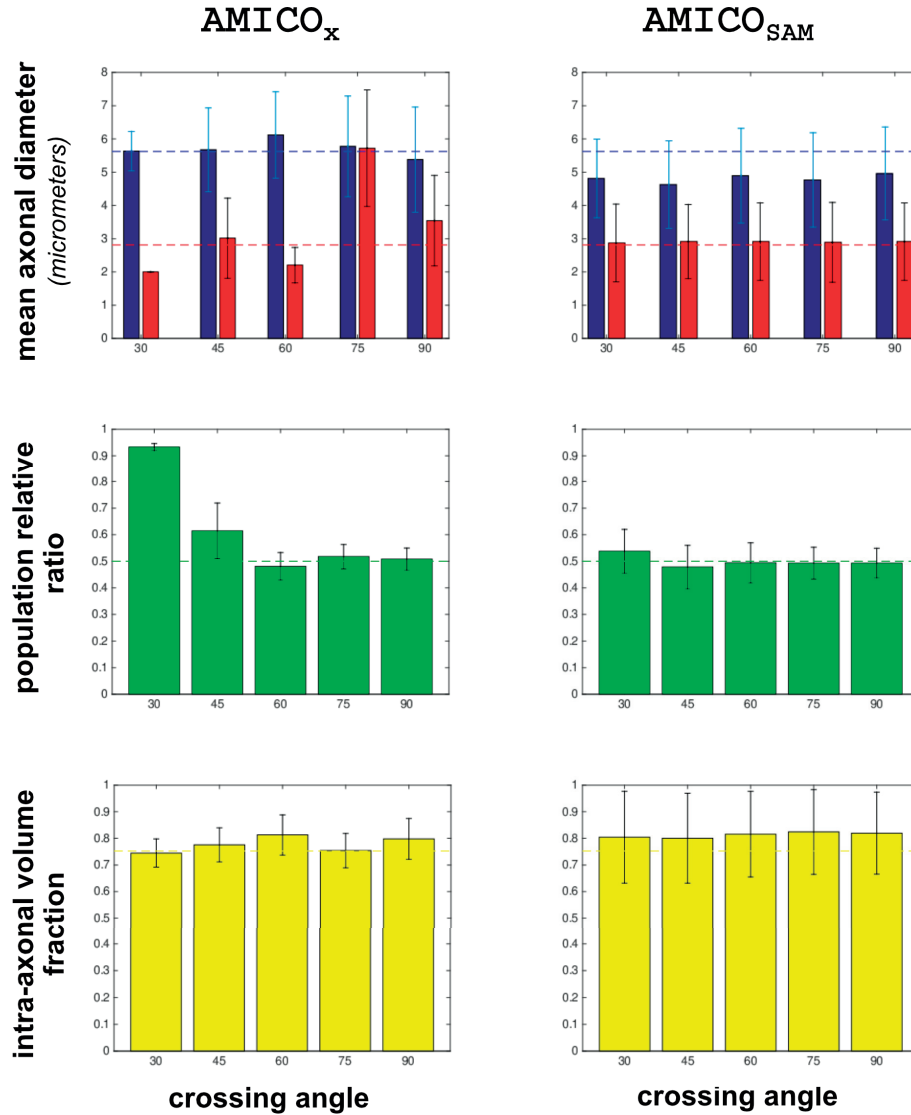


Figure 6.10: Detailed comparison of the mean axonal diameter index and population relative ratio estimated with AMICO_x (left) and AMICO_{SAM} (right), as a function of the crossing angle between the fiber populations. Results correspond to an undersampled protocol with **54 measurements**, corresponding to $\approx 20\%$ of the fully-sampled protocol.

Study different q -space undersampling strategies

AMICO_{SAM} appears as an extension of the AMICO framework to enable microstructure reconstructions from a reduced number of measurements, thus leading to faster acquisitions. To undersample our fully-sampled dataset, so far we have uniformly reduced the number of points in every shell, i. e. our undersampled datasets contain the same number of data points in every shell and they are located along the same orientations. Different undersampling strategies should be investigated. As pointed out in [106], a careful design is central in the success of multi-shell acquisition and reconstruction techniques. In their

work, the authors provide a general method to design multi-shell acquisition with uniform angular coverage that should be studied in the framework of our problem.

Study different regularization strategies

As mentioned in section 6.3.1, the weights regularize the coefficients and thus, the proposed reweighting scheme imposes smoothness through structured sparsity of the solution. Thanks to the flexibility of convex optimization, different prior information can be easily imposed through a different definition of the weights and its effect can therefore be investigated. For instance, we could make only the intra-axonal compartments be dependent on the surrounding voxels and not the extra-axonal ones, or vice versa.

An alternative regularization function, a weighted mixed $\ell_{2,1}$ norm[†] to promote *group sparsity* in the solution was also considered. Preliminary results justify a deeper investigation in that direction. Moreover, the reformulation of the unconstrained problem into a constrained one would avoid tuning the regularization parameter λ , provided a good approximation of the overall bound on the level of the noise could be estimated.

Further validation

Synthetic data obtained from a Monte-Carlo diffusion simulator offers a good starting validation strategy, commonly used by the community [1, 9]. However, a validation of AMICO_{SAM} on more complex phantoms to test its ability to correctly assess the number and orientation of the fiber populations, as well as their microstructure features in a more challenging context appears as the next natural step. Furthermore, other validation strategies both on *ex-vivo* and *in-vivo* data should be designed and obviously follow.

6.4 Conclusion

In this chapter, we have first presented AMICO_X, a convex framework for microstructure imaging that enables fast microstructure feature mapping in regions with multiple fiber populations. In the second part, we have applied the idea of spatial regularization through structured sparsity described in chapter 5 to the microstructure recovery problem. As a result, we propose a new formulation that estimates simultaneously the orientation and the microstructure of the fiber bundles. AMICO_{SAM} exploits neighboring information to impose a strong regularization in the so-called “*spatial-angular-microstructure*” (SAM) space. Consequently, it enables robust reconstructions from a low number of measurements, thus leading to faster acquisitions.

Preliminary results on synthetic data show that AMICO_{SAM} has the potential to disentangle different fiber populations and to correctly assess both intra-axonal volume fraction

[†]For an arbitrary matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ its $\ell_{2,1}$ norm is defined as $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^m \sqrt{\sum_{j=1}^n x_{ij}^2}$.

and relative population ratio with as few as ≈ 50 measurements (as typical HARDI protocols). It is plausible to think that the powerful regularization prior in the SAM space can as well guarantee a strong robustness to noise. $\text{AMICO}_{\text{SAM}}$ should be tested under different noise conditions to confirm this hypothesis. Also, further validation on real data is still needed.

We highlight that $\text{AMICO}_{\text{SAM}}$ provides a versatile and flexible model-independent framework for microstructure imaging. In this chapter we have presented results on $\text{AMICO}_{\text{SAM}}$ for ActiveAx [1], but, as the original AMICO [9], it can potentially be applied to any other microstructure imaging modality. Moreover, the weights can be easily redefined to impose smoothness priors according to the context of study.

The fact that $\text{AMICO}_{\text{SAM}}$ simultaneously estimates the fiber orientations and their microstructure features, brings this algorithm specially close to the field of *microstructure informed tractography* [107, 108], that combines microstructure information and tractography. Future work should consider merging the principles behind $\text{AMICO}_{\text{SAM}}$ with algorithms such as COMMIT [107].

Generalization to multiple correlated sparse signal recovery

Before concluding, we would like to highlight that our proposed formulation to promote spatial regularization through a structured sparsity prior can actually be applied in a more generic framework for *multiple correlated sparse signal recovery*. In sections 5.2 and 6.3.1, spatially-correlated vectors are concatenated to build matrix \mathbf{X} . Formulations (5.5) and (6.5) can actually be generalized to recover multiple sparse signals correlated through a smoothness prior on the variation of the signal support in the inter-signal dimension. Indeed, considering the concatenation of n_D *correlated* sparse signals into a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$ and a linear operator Φ that models a measurement process on them, equation (5.5) can be rewritten as:

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}} \|\Phi(\mathcal{X}) - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathcal{X}\|_{\mathcal{W},1} \leq K, \quad (6.8)$$

with $\|\cdot\|_{\mathcal{W},1}$ denoting a weighted- ℓ_1 norm of a generic tensor defined as:

$$\|\mathcal{X}\|_{\mathcal{W},1} = \sum_{i_1, \dots, i_D} \mathcal{W}_{i_1 \dots i_D} |\mathcal{X}_{i_1 \dots i_D}|. \quad (6.9)$$

The definition of the weights and neighborhoods will enable the embedding of the smoothness prior in the signal support through the structure on the sparsity and must, of course, be adapted to the application. Our method stands in contrast to other joint sparsity models [109, 110] that assume a common support of the correlated signals. Social sparsity models [111, 112] also leverage the concept of neighborhoods to promote sparsity. However, our proposed algorithms are essentially inspired from the reweighting scheme

proposed by Candes et al. [24] to approach ℓ_0 minimization through the convex minimization of a weighted ℓ_1 norm. Our contributions lie in enforcing a *structured version* of the ℓ_0 norm of the solution at convergence and thus, the weights should continue to represent the inverse value of the associated entry at the previous iteration, to converge to an ℓ_0 norm.

Part of the work presented in this chapter has been published in [8, 10, 11, 12, 13].

CHAPTER 7

Conclusion

In this thesis, we have studied sparse Fourier sensing problems in the fields of optical interferometry and diffusion MRI. In both applications, we face ill-posed (with less available measurements than the dimension of the signal) and highly nonlinear problems. Our strategy to solve these originally nonlinear problems resorts to reformulating them as linear inverse problems that can be solved using convex optimization methods. Inspired by the compressed sensing framework, we propose novel priors to leverage the sparse structure (i.e. low dimensionality) of the solution.

All nice properties of the convex optimization framework prove to be the cornerstone of our contributions:

- *From the formulation point of view*, convex optimization constitutes an extremely **adaptable framework** that enables the straightforward inclusion of prior information about the signal, as long as it is formulated as a convex constraint. In this thesis we have immensely taken advantage of this characteristic to design the proposed novel algorithms. In the application for optical interferometry, we regularize the linearized inverse problem through a nuclear norm relaxation of a low rank constraint, easily imposing also positive-semidefiniteness, reality, positivity and optionally sparsity constraints. In diffusion MRI, we resort to a reweighted scheme and design a flexible weighting system that enforces structured sparsity and promotes spatial regularization in the solution simultaneously. Thanks to this flexibility, we design powerful regularization priors that guarantee a strong robustness to noise and the ability to go to higher undersampling regimes. In diffusion MRI, where the measuring process is considerably time consuming, this ability of solving the problem with few data points appears crucial for the technique to be clinically feasible.
- *From the reconstruction point of view*, convex problems **converge to a global minimum** and **do not depend on the initialization**. These two properties are key to our contribution in the field of optical interferometry, where state-of-the-art methods generally perform local optimization, highly dependent on both the initial image and the path followed by the method. Moreover, convex problems can conveniently be solved using **fast proximal splitting methods**. The increase of speed with respect to other nonlinear routines represents also a contribution in the

field of microstructure imaging.

We note that our original formulation presented in Part I can be seen as a generalization of the Phase Lift approach for the phase retrieval problem [2]. Indeed, we lift the ambient dimension of the problem to formulate a tensor completion problem from a set of linear measurements on the tensor. We have therefore proposed a novel linear alternative to deal with bispectrum (nonlinear) measurements that can be useful to other fields facing bispectral analysis.

We would also like to highlight that the method presented in Part II of this thesis to promote spatial regularization through a structured sparsity prior can actually be applied in a more generic framework for *multiple correlated sparse signal recovery*. A generalization of our formulation can account for multidimensional signals and correlations of different nature - such as temporal - between them. The definition of the neighborhoods as well as the weights should be adapted to the application.

7.1 Perspectives

7.1.1 Optical Interferometry

Among our two proposed methods to solve the imaging problem in the field of optical interferometry, the AM approach has been further studied and developed in [63], including sparsity priors and presenting convergence guarantees. On the other hand, our proposed linear and convex approach NM ensures desirable properties of convergence and independence on the initialization and we are convinced it is worth a deeper study. The main challenge associated to the practical implementation of this approach lies in the increase of the ambient dimension, as the price to pay for the linearization. Indeed, the linear approaches NM and NM-RW turn out to be extremely demanding from the computational point of view. Therefore, in our opinion, all efforts should be focused on studying software and hardware optimization to solve the problem for higher dimension images, e.g. using graphics processing units [61]. Additionally, recent approaches for radio interferometry [60] justify the investigation of different kind of sparsity priors, thanks to the versatility of convex optimization. Finally, our theoretical results should be confirmed on real data and be explicitly compared to other state-of-the-art methods. Bringing our theoretical results closer to reality would enhance our contribution that remains a unique approach to formulate the challenging problem of image reconstruction in optical interferometry as a tensor completion problem.

7.1.2 diffusion MRI

Unified framework with fiber tracking

Results of the HARDI reconstruction challenge (ISBI 2013) [113], in which we participated, already gave evidence that local reconstruction methods and global tractography cannot be treated separately any more. In our opinion, intra-voxel dMRI and global tractography should be unified in a single framework, intending to (i) reconstruct the fiber configuration (orientation and microstructure) in each voxel and (ii) solve the fiber-tracking across the whole brain, simultaneously. Work in this direction should consider merging our approach with COMMIT [107], that provides a Convex Optimization Model for Microstructure Informed Tractography.

Combined k -space and q -space undersampling

Further evolutions of our method for fiber reconstruction (either orientation or microstructure) should consider undersampling in Fourier space (k -space) for each of the q -space images acquired, along the lines of work in [86]. Indeed, the assumption that our function of interest is sparse in fiber space enables q -space undersampling, as we have shown. As soon as the problem is formulated globally (i.e., solved for all voxels simultaneously instead of voxelwise), as we do, the sparsity of the signal can also be considered in voxel space; thus leading to combined q -space and k -space undersampling and potentially resulting into a significant additional acceleration.

In a setting where the acquisition can be accelerated either by undersampling in q -space (probing the signal with less diffusion vectors) and/or in k -space (probing the signal in less voxels), the optimal acceleration scheme in terms of acquisition speed (number of measurements) and reconstruction quality should be studied. Recent work [114, 115] initiated this line of investigation for fiber orientation recovery; analogous work for microstructure estimation should follow.

Implementation on real MR scanners

In close relationship with our previous point, implementation on real MR scanners of sampling sequences optimized for our proposed formulation should follow. Collecting real data and confirming our numerical results would represent the ultimate validation of the ability of our novel formulation to perform fiber reconstruction in all regions of the brain in a clinical context.

Bibliography

- [1] D. C. Alexander, P. L. Hubbard, M. G. Hall, E. A. Moore, M. Ptito, G. J. M. Parker, and T. B. Dyrby, “Orientationally invariant indices of axon diameter and density from diffusion MRI,” *NeuroImage*, vol. 52, no. 4, pp. 1374–1389, 2010.
- [2] E. Candès, T. Strohmer, and V. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Communications on Pure and Applied Mathematics*, 2011, preprint in arXiv:1109.4499v1.
- [3] A. Auría, R. E. Carrillo, J.-P. Thiran, and Y. Wiaux, “Tensor optimization for optical-interferometric imaging,” *MNRAS*, vol. 437, no. 3, pp. 2083–2091, 2014.
- [4] A. Auría, R. E. Carrillo, J.-P. Thiran, and Y. Wiaux, “Sparsity in tensor optimization for optical-interferometric imaging,” in *2014 IEEE International Conference on Image Processing*. IEEE, 2014, pp. 6026–6030.
- [5] J. Tournier, F. Calamante, D. Gadian, and A. Connelly, “Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution,” *NeuroImage*, vol. 23, pp. 1176–1185, 2004.
- [6] A. Ramirez-Manzanares, M. Rivera, B. Vemuri, P. Carney, and T. Mareci, “Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry,” *IEEE Transactions on Medical Imaging*, vol. 26, pp. 1091–1102, 2007.
- [7] A. Auría, A. Daducci, J.-P. Thiran, and Y. Wiaux, “Structured sparsity for spatially coherent fibre orientation estimation in diffusion MRI,” *NeuroImage*, vol. 115, pp. 245 – 255, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1053811915003493>
- [8] A. Auría, A. Daducci, J.-P. Thiran, and Y. Wiaux, “Structured sparsity through reweighting and application to diffusion MRI,” in *23rd European Signal Processing Conference*, 2015.
- [9] A. Daducci, E. J. Canales-Rodriguez, H. Zhang, T. B. Dyrby, D. C. Alexander, and J.-P. Thiran, “Accelerated Microstructure Imaging via Convex Optimization

- (AMICO) from diffusion MRI data,” *Neuroimage*, vol. 105, pp. 32–44, 2015. [Online]. Available: <https://github.com/daducci/AMICO>
- [10] A. Auría, E. Canales-Rodriguez, Y. Wiaux, T. Dyrby, D. Alexander, J.-P. Thiran, and A. Daducci, “Accelerated Microstructure Imaging via Convex Optimization (AMICO) in crossing fibers,” in *23rd annual meeting of the International Society for Magnetic Resonance in Medicine (ISMRM)*, 2015.
 - [11] A. Auría, D. P. R. Romascano, E. Canales-Rodriguez, Y. Wiaux, T. B. Dirby, D. Alexander, J.-P. Thiran, and A. Daducci, “Accelerated Microstructure Imaging via Convex Optimisation for regions with multiple fibres (AMICOx),” in *IEEE International Conference on Image Processing 2015*, 2015.
 - [12] A. Auría, E. Canales-Rodriguez, Y. Wiaux, T. Dyrby, D. Alexander, J.-P. Thiran, and A. Daducci, “Reducing acquisition time for axon diameter mapping using global optimization in the spatial-angular-microstructure space,” in *24th annual meeting of the International Society for Magnetic Resonance in Medicine (ISMRM)*, 2016.
 - [13] A. Auría, D. P. R. Romascano, E. J. Canales-Rodriguez, T. B. Dyrby, D. C. Alexander, J.-P. Thiran, Y. Wiaux, and A. Daducci, “Reducing acquisition time for microstructure imaging with spatially-regularized global optimization,” in *22nd Annual Meeting of the Organization for Human Brain Mapping (OHBM)*, 2016.
 - [14] J. Hastad, “Tensor rank is NP-complete,” *Journal of Algorithms*, vol. 11, pp. 644–654, 1990.
 - [15] G. W. Stewart, “On the early history of the singular value decomposition,” *SIAM Review*, vol. 35, no. 4, pp. 551–566, 1992.
 - [16] T. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, pp. 455–500, 2009.
 - [17] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.
 - [18] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.
 - [19] E. J. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec 2006.
 - [20] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

-
- [21] M. Fornasier and H. Rauhut, *Handbook of Mathematical Methods in Imaging*. Springer, 2011, ch. Compressed sensing.
 - [22] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012, ch. Introduction to Compressed Sensing.
 - [23] E. J. Candès, Y. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2010.
 - [24] E. Candès, M. Wakin, and S. Boyd, “Enhancing sparsity by weighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.
 - [25] E. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, pp. 717–762, 2009.
 - [26] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inf. Theor.*, vol. 56, no. 5, pp. 2053–2080, May 2010.
 - [27] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
 - [28] E. J. Candès, T. Strohmer, and V. Voroninski, “Phase lift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2011.
 - [29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Cambridge UK, 2004.
 - [30] P. L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” *Fixed-Point Algorithms for Inverse Problems in Science and Engineering (Bauschke H. H., Burachik R.S., Combettes P.L., Elser V., Luke D.R., Wolkowicz H., eds)*, pp. 185–212, 2011.
 - [31] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, “Purify: a new approach to radio-interferometric imaging,” *MNRAS*, vol. 439, no. 4, pp. 3591–3604, 2014.
 - [32] E. Thiébaud and J. Giovanelli, “Image reconstruction in optical interferometry,” *IEEE Signal Processing Magazine*, vol. 21, pp. 97–109, 2010.
 - [33] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*. Wiley-Interscience, New York, 2001.

-
- [34] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vandergheynst, “Compressed sensing imaging techniques for radio interferometry,” *MNRAS*, vol. 395, no. 3, pp. 1733–1742, 2009.
 - [35] Y. Wiaux, G. Puy, Y. Boursier, and P. Vandergheynst, “Spread spectrum for imaging techniques in radio interferometry,” *MNRAS*, vol. 400, no. 2, pp. 1029–1038, 2009.
 - [36] Y. Wiaux, G. Puy, and P. Vandergheynst, “Compressed sensing reconstruction of a string signal from interferometric observations of the cosmic microwave background,” *MNRAS*, vol. 402, no. 4, pp. 2626–2636, 2010.
 - [37] J. D. McEwen and Y. Wiaux, “Compressed sensing for wide-field radio interferometric imaging,” *MNRAS*, vol. 413, no. 2, pp. 1318–1332, 2011.
 - [38] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, “Sparsity averaging reweighted analysis (sara): a novel algorithm for radio-interferometric imaging,” *MNRAS*, vol. 426, no. 2, pp. 1223–1234, 2012.
 - [39] F. Li, T. J. Cornwell, and F. de Hoog, “Application of compressive sampling to radio astronomy I: Deconvolution,” *A&A*, vol. A31, pp. 528–538, 2011.
 - [40] S. Wenger, M. Magnor, Y. Pihlströsm, S. Bhatnagar, and U. Rau, “Sparseri: A compressed sensing framework for aperture synthesis imaging in radio astronomy,” *Publications of the Astronomical Society of the Pacific*, vol. 122, no. 897, pp. 1367–1374, 2010.
 - [41] H. Thorsteinsson, D. F. Buscher, and J. S. Young, “The bispectrum in model-independent imaging,” *Proceedings of SPIE, New Frontiers in Stellar Interferometry*, vol. 5491, pp. 1574–1579, 2004.
 - [42] J. Baldwin and C. Haniff, “The application of interferometry to optical astronomical imaging,” *Trans. R. Soc. London*, vol. A 360, pp. 969–986, 2002.
 - [43] J. D. Monnier, “Optical interferometry in astronomy,” *Reports on Progress in Physics*, vol. 66, pp. 789–857, May 2003.
 - [44] J. Baldwin, C. Haniff, C. Mackay, and P. Warner, “Closure phase in high-resolution optical imaging,” *Nature*, vol. 320, pp. 595–597, 1986.
 - [45] C. A. Haniff, C. D. Mackay, D. J. Titterton, D. Sivia, J. E. Baldwin, and P. J. Warner, “The first images from optical aperture synthesis,” *Nature*, vol. 328, pp. 694–696, 1987.
 - [46] P. G. Tuthill, J. D. Monnier, W. C. Danchi, E. H. Wishnow, and C. A. Haniff, “Michelson interferometry with the Keck I Telescope,” *Publications of the Astronomical Society of the Pacific*, vol. 112, no. 770, pp. 555–565, Apr. 2000.

-
- [47] E. Thiébaud, “MIRA: an effective imaging algorithm for optical interferometry,” *Proceedings of SPIE, Optical and Infrared Interferometry*, vol. 7013, p. 70131I, 2008.
 - [48] S. Renard, E. Thiébaud, and F. Malbet, “Image reconstruction in optical interferometry: Benchmarking the regularization,” *A&A*, vol. 533, p. A64, 2011.
 - [49] S. C. Meimon, L. M. Mugnier, and G. L. Besnerais, “Reconstruction method for weak-phase optical interferometry,” *Opt. Lett.*, vol. 30, no. 14, pp. 1809–1811, Jul 2005.
 - [50] H. Attouch, J. Bolte, and B. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods,” *Mathematical Programming*, vol. 137, pp. 91–129, 2013.
 - [51] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, “Proximal alternating minimization and projection methods for nonconvex problems,” *Mathematics of Operations Research*, vol. 35, no. 3, pp. 438–457, 2010.
 - [52] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
 - [53] J. Haldar and D. Hernando, “Rank-constrained solutions to linear matrix equations using powerfactorization,” *IEEE Signal Processing Letters*, vol. 16, pp. 584–587, 2009.
 - [54] S. Gandy, B. Recht, and I. Yamada, “Tensor completion and low-n-rank tensor recovery via convex optimization,” *Inverse Problems*, vol. 27, p. 025010, 2011.
 - [55] E. Kofidis and P. A. Regalia, “On the best rank-1 approximation of higher-order supersymmetric tensors,” *SIAM J. Matrix Anal. Appl.*, vol. 23, pp. 863–884, 2002.
 - [56] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1989.
 - [57] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal of Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
 - [58] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.
 - [59] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, “Phase retrieval via matrix completion,” *SIAM J. on Imaging Sciences*, vol. 6, no. 1, pp. 199–225, 2011.

-
- [60] R. E. Carrillo, J. D. McEwen, D. VanDeVille, J.-P. Thiran, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Processing Letters*, vol. 20, no. 6, pp. 591–594, 2013.
 - [61] F. Baron and B. Kloppenborg, "GPU-accelerated image reconstruction for optical and infrared interferometry," *Proceedings of SPIE, Optical and Infrared Interferometry II*, vol. 7734, p. 77344D, 2010.
 - [62] J. Ranieri, A. Chebira, Y. M. Lu, and M. Vetterli, "Phase retrieval for sparse signals: Uniqueness conditions," *CoRR*, vol. abs/1308.3058, 2013.
 - [63] J. Birdi, A. Repetti, and Y. Wiaux, "A regularized tri-linear approach for optical interferometric imaging," 2016, submitted to MNRAS. Preprint available at <https://arxiv.org/abs/1609.00546>.
 - [64] C. Beaulieu, "The basis of anisotropic water diffusion in the nervous system – a technical review," *NMR in Biomedicine*, vol. 15, no. 7-8, pp. 435–455, 2002.
 - [65] E. Stejskal and J. Tanner, "Spin diffusion measurements - spin echoes in presence of a time-dependent field gradient," *J Chem Phys*, vol. 42, pp. 288–300, 1965.
 - [66] D. M. Patterson, A. R. Padhani, and D. J. Collins, "Technology insight: water diffusion MRI – a potential new biomarker of response to cancer therapy," *Nat Clin Prac Oncol*, vol. 5, no. 4, pp. 220–233, 04 2008. [Online]. Available: <http://dx.doi.org/10.1038/ncponc1073>
 - [67] P. Callaghan, *Principles of nuclear magnetic resonance microscopy*. Clarendon Press, Oxford, 1991.
 - [68] P. Hagmann, L. Jonasson, P. Maeder, J.-P. Thiran, V. J. Wedeen, and R. Meuli, "Understanding diffusion mr imaging techniques: From scalar diffusion-weighted imaging to diffusion tensor imaging and beyond," *RadioGraphics*, vol. 26, no. suppl_1, pp. S205–S223, 2006.
 - [69] J. Ritchie, "On the relation between fibre diameter and conduction velocity in myelinated nerve fibres," *Proceedings of the Royal Society of London, Series B Biological Sciences*, vol. 27, pp. 29–35, 1982.
 - [70] M. Shintaku, A. Hirano, and J. Llana, "Increased diameter of demyelinated axons in chronic multiple sclerosis of the spinal cord," *Neuropathol. Appl. Neurobiol.*, vol. 14, pp. 505–510, 1982.
 - [71] P. Basser, J. Mattiello, and D. Le Bihan, "MR diffusion tensor spectroscopy and imaging," *Biophysical journal*, vol. 66, pp. 259–267, 1994.

-
- [72] D. Alexander, "Maximum entropy spherical deconvolution for diffusion MRI," *Information Processing in Medical Imaging (IPMI)*, pp. 76–87, 2005.
- [73] J. Tournier, F. Calamante, and A. Connelly, "Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution," *NeuroImage*, vol. 35, pp. 1459–1472, 2007.
- [74] D. Barazany, P. Basser, and Y. Assaf, "In vivo measurement of axon diameter distribution in the corpus callosum of rat brain," *Brain*, vol. 132, no. 5, pp. 1210–1220, 2009.
- [75] E. Panagiotaki, T. Schneider, B. Siow, M. G. Hall, M. F. Lythgoe, and D. C. Alexander, "Compartment models of the diffusion MR signal in brain white matter: A taxonomy and comparison," *NeuroImage*, vol. 59, no. 3, pp. 2241–2254, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.neuroimage.2011.09.081>
- [76] A. Szafer, J. Zhong, and J. C. Gore, "Theoretical model for water diffusion in tissues," *Magnetic Resonance in Medicine*, vol. 33, no. 5, pp. 697–712, 1995. [Online]. Available: <http://dx.doi.org/10.1002/mrm.1910330516>
- [77] T. Behrens, M. Woolrich, M. Jenkinson, H. Johansen-Berg, R. Nunes, S. Clare, P. Matthews, J. Brady, and S. Smith, "Characterization and propagation of uncertainty in diffusion-weighted MR imaging," *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 1077–1088, 2003. [Online]. Available: <http://dx.doi.org/10.1002/mrm.10609>
- [78] Y. Assaf, R. Z. Freidlin, G. K. Rohde, and P. J. Basser, "New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter," *Magnetic Resonance in Medicine*, vol. 52, no. 5, pp. 965–978, 2004. [Online]. Available: <http://dx.doi.org/10.1002/mrm.20274>
- [79] Y. Assaf and P. J. Basser, "Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain," *NeuroImage*, vol. 27, no. 1, pp. 48–58, 2005.
- [80] Y. Assaf, T. Blumenfeld-Katzir, Y. Yovel, and P. J. Basser, "Axciliber: A method for measuring axon diameter distribution from diffusion MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 6, pp. 1347–1354, 2008. [Online]. Available: <http://dx.doi.org/10.1002/mrm.21577>
- [81] D. C. Alexander, "A general framework for experiment design in diffusion mri and its application in measuring direct tissue-microstructure features," *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 439–448, 2008. [Online]. Available: <http://dx.doi.org/10.1002/mrm.21646>

-
- [82] H. Zhang, T. Schneider, C. A. M. Wheeler-Kingshott, and D. C. Alexander, "NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain." *NeuroImage*, vol. 61, no. 4, pp. 1000–1016, 2012.
 - [83] G. Stainsz, A. Szafer, G. Wright, and R. Henkelman, "An analytical model of restricted diffusion in bovine optic nerve," *Magnetic Resonance in Medicine*, vol. 37, no. 1, pp. 103–111, 1997.
 - [84] V. J. Wedeen, P. Hagmann, W. Tseng, T. Reese, and R. Weisskoff, "Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 54, pp. 1377–1386, 2005.
 - [85] B. Jian and B. Vermuri, "A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 1464–1471, 2007.
 - [86] M. Mani, M. Jacob, A. Guidon, V. Magnotta, and J. Zhong, "Acceleration of high angular and spatial resolution diffusion imaging using compressed sensing with multichannel spiral data," *Magnetic Resonance in Medicine*, 2014.
 - [87] A. Daducci, D. V. D. Ville, J. Thiran, and Y. Wiaux, "Sparse regularization for fiber ODF reconstruction: from the suboptimality of ℓ_2 and ℓ_1 priors to ℓ_0 ," *Medical Image Analysis*, vol. 18, no. 6, pp. 820–33, 2014.
 - [88] A. Tristán-Vega and S. Aja-Fernández, "DWI filtering using joint information for DTI and HARDI." *Medical Image Analysis*, vol. 14, no. 2, pp. 205–218, 2010.
 - [89] N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot, "Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI." ser. Lecture Notes in Computer Science, vol. 5242. Springer, 2008, pp. 171–179.
 - [90] P. Fillard, X. Pennec, V. Arsigny, and N. Ayache, "Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics," *IEEE Transactions on Medical Imaging*, vol. 26, no. 11, pp. 1472–1482, November 2007. [Online]. Available: <http://www.inria.fr/sophia/asclepios/Publications/Pierre.Fillard/Fillard.TMI.2007.pdf>
 - [91] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992. [Online]. Available: [http://dx.doi.org/10.1016/0167-2789\(92\)90242-F](http://dx.doi.org/10.1016/0167-2789(92)90242-F)
 - [92] O. Michailovich, Y. Rathi, and S. Dolui, "Spatially regularized compressed sensing for high angular resolution diffusion imaging," *IEEE Transactions on Medical Imaging*, vol. 30, pp. 1100–1115, 2011.

-
- [93] J. Huang, R. Yagel, V. Filippov, and Y. Kurzion, "An accurate method for voxelizing polygon meshes," *Proceedings of the 1998 IEEE Symposium on Volume Visualization*, pp. 119–126, 1998.
- [94] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, 2006.
- [95] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.
- [96] A. Daducci, E. Canales-Rodriguez, M. Descoteaux, E. Garyfallidis, Y. Gur, Y. Lin, M. Mani, S. Merlet, M. Paquette, A. Ramirez-Manzanares, M. Reisert, P. Rodrigues, F. Seppehrband, E. Caruyer, J. Choupan, R. Deriche, M. Jacob, G. Menegaz, V. Prökovska, M. Rivera, Y. Wiaux, and J. Thiran, "Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 384–399, 2014.
- [97] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, 1995.
- [98] D. Jones, M. Horsfield, and A. Simmons, "Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 42, pp. 515–525, 1999.
- [99] F. Yeh and W. Tseng, "Sparse solution of fiber orientation distribution function by diffusion decomposition," *PLoS One*, vol. 8, 2013.
- [100] V. Prcökovska, P. Rodrigues, R. Duits, B. ter Haar Romeny, and A. Vilanova, "Extrapolating fiber crossings from DTI data. Can we gain similar information as HARDI?" *Proc. Workshop computational diffusion MRI, MICCAI*, pp. 26–37, 2010.
- [101] E. Özarslan, T. Shepherd, B. Vemuri, S. Blackband, and T. Mareci, "Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT)," *NeuroImage*, vol. 31, pp. 1086–1103, 2006.
- [102] E. Canales-Rodriguez, Y. Iturria-Medina, Y. Alemaán-Gómez, and L. Melie-Garcia, "Deconvolution in diffusion spectrum imaging," *NeuroImage*, vol. 50, pp. 136–149, 2010.
- [103] T. Dyrby, L. V. Søgaard, M. G. Hall, M. Ptito, and D. Alexander, "Contrast and stability of the axon diameter index from microstructure imaging with diffusion MRI," *Magnetic Resonance in Medicine*, vol. 70, no. 3, pp. 711–721, 2013. [Online]. Available: <http://dx.doi.org/10.1002/mrm.24501>

-
- [104] H. Zhang, T. Dyrby, and D. Alexander, “Axon Diameter Mapping in Crossing Fibers with Diffusion MRI.” ser. Lecture Notes in Computer Science, vol. 6892. Springer, 2011, pp. 82–89.
 - [105] M. G. Hall and D. C. Alexander, “Convergence and Parameter Choice for Monte-Carlo Simulations of Diffusion MRI.” *IEEE Trans. Med. Imaging*, vol. 28, no. 9, pp. 1354–1364, 2009. [Online]. Available: <http://dblp.uni-trier.de/db/journals/tmi/tmi28.html#HallA09>
 - [106] E. Caruyer, C. Lenglet, G. Sapiro, and R. Deriche, “Design of multishell sampling schemes with uniform coverage in diffusion mri,” *Magnetic Resonance in Medicine*, vol. 69, no. 6, pp. 1534–1540, 2013. [Online]. Available: <http://dx.doi.org/10.1002/mrm.24736>
 - [107] A. Daducci, A. D. Palù, A. Lemkaddem, and J. P. Thiran, “COMMIT: Convex optimization modeling for microstructure informed tractography,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 246–257, Jan 2015.
 - [108] A. Daducci, A. Dal Palù, M. Descoteaux, and J.-P. Thiran, “Microstructure informed tractography: Pitfalls and open challenges,” *Frontiers in Neuroscience*, vol. 10, p. 247, 2016.
 - [109] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, “Distributed compressed sensing of jointly sparse signals,” in *In Asilomar Conf. Signals, Sys., Comput*, 2005, pp. 1537–1541.
 - [110] M. F. Duarte and Y. C. Eldar, “Structured compressed sensing: From theory to applications,” *IEEE TRANS. SIGNAL PROCESS*, vol. 59, no. 9, pp. 4053–4085, 2011.
 - [111] M. Kowalski and B. Torrèsani, “Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients.” *Signal, Image and Video Processing*, vol. 3, no. 3, pp. 251–264, 2009.
 - [112] M. Kowalski, K. Siedenburg, and M. Dörfler, “Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators.” *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2498–2511, 2013. [Online]. Available: <http://dblp.uni-trier.de/db/journals/tsp/tsp61.html#KowalskiSD13>
 - [113] http://hardi.epfl.ch/static/events/2013_ISBI/, 2013.
 - [114] M. Pesce, J.-P. Thiran, and Y. Wiaux, “Joint kq-space acceleration for fibre orientation mapping in diffusion MRI,” in *22nd Annual Meeting of the Organization for Human Brain Mapping (OHBM)*, 1 2016.

-
- [115] M. Pesce, A. Auría, A. Daducci, J.-P. Thiran, and Y. Wiaux, “Joint kq-space acceleration for fibre orientation estimation in diffusion MRI,” in *International Biomedical and Astronomical Signal Processing Frontiers Workshop, BASP*, 1 2017.

List of Publications

Journal Publications

- 1 **A. Auría**, A. Daducci, J.-P. Thiran and Y. Wiaux, "Structured sparsity for spatially coherent fibre orientation estimation in diffusion MRI", *NeuroImage*, vol. 115, pp. 245-255, 2015.
- 2 **A. Auría**, R. E. Carrillo, J.-P. Thiran and Y. Wiaux, "Tensor optimization for optical-interferometric imaging", *MNRAS*, vol. 437, no.3, pp. 2083-2091, 2014.

Conference Publications

- 1 M. Pesce, **A. Auría**, A. Daducci, J.-P. Thiran and Y. Wiaux, "Joint kq-space acceleration for fibre orientation estimation in diffusion MRI". *International BASP Frontiers workshop*, 2017.
- 2 D. P. Romascano, M. Barakovic, **A. Auría**, T.B. Dirby, J.-P. Thiran and A. Daducci, "Orientation invariant and non-parametric Axon Diameter Distribution mapping using PGSE and regularized discrete linear modeling". accepted to *the 25th annual meeting of the International Society for Magnetic Resonances in Medicine (ISMRM)*, 2017.
- 3 **A. Auría**, D. P. Romascano, E. Canales-Rodriguez, Y. Wiaux, T.B. Dirby, D. Alexander, J.-P. Thiran and A. Daducci, "Reducing acquisition time for microstructure imaging with spatially-regularized global optimization". in *the 22nd Annual Meeting of the Organization for Human Brain Mapping (OHBM)*, 2016.
- 4 **A. Auría**, E. Canales-Rodriguez, Y. Wiaux, T.B. Dirby, D. Alexander, J.-P. Thiran and A. Daducci, "Reducing acquisition time for axon diameter mapping using global optimization in the spatial-angular-microstructure space". in *the 24th annual meeting of the International Society for Magnetic Resonances in Medicine (ISMRM)*, 2016.

-
- 5 **A. Auría**, A. Daducci, J.-P. Thiran and Y. Wiaux, "Structured sparsity through reweighting and application to diffusion MRI", in *23rd European Signal Processing Conference*, 2015.
 - 6 **A. Auría**, D.P. Romascano, E. Canales-Rodriguez, Y. Wiaux, T.B. Dirby, D. Alexander, J.-P. Thiran and A. Daducci, "Accelerated Microstructure Imaging via Convex Optimisation for regions with multiple fibres (AMICOx)". *IEEE International Conference on Image Processing, ICIP*, IEEE 2015.
 - 7 **A. Auría**, E. Canales-Rodriguez, Y. Wiaux, T.B. Dirby, D. Alexander, J.-P. Thiran and A. Daducci, "Accelerated Microstructure Imaging via Convex Optimization (AMICO) in crossing fibers". in *the 23rd annual meeting of the International Society for Magnetic Resonances in Medicine (ISMRM)* , 2015.
 - 8 **A. Auría**, R. E. Carrillo, J.-P. Thiran and Y. Wiaux, "Sparsity in tensor optimization for optical-interferometric imaging". *IEEE International Conference on Image Processing, ICIP*, IEEE 2014, pp. 6026-6030.
 - 9 **A. Auría**, R. E. Carrillo and Y. Wiaux, "A convex optimization approach for image recovery from nonlinear measurements in optical interferometry". *International BASP Frontiers workshop*, 2013.
 - 10 A. Daducci, **A. Auría**, J.-P. Thiran and Y. Wiaux, " ℓ_0 -deconvolution for compressive diffusion MRI". *International BASP Frontiers workshop*, 2013

Anna Auría

CONTACT INFORMATION	EPFL STI IEL LTS5 ELD 241 (Station 11) 1015, Lausanne, Switzerland	☎ +41 21 693 69 33 ✉ anna.auria@epfl.ch
CITIZENSHIP	Spanish	
RESEARCH INTERESTS	inverse problems, compressed sensing, sparsity, structured sparsity, convex optimization, diffusion MRI, microstructure imaging.	
EDUCATION	École Polytechnique Fédérale de Lausanne, EPFL PhD candidate Electrical and Electronics Engineering, Signal Processing Laboratory 5 LTS5	<i>2011-now</i>
	Universitat Politècnica de Catalunya, UPC Double Diploma: Master in Mathematics and Electrical Engineering	<i>2000-2007</i>
PROFESSIONAL EXPERIENCE	École Polytechnique Fédérale de Lausanne, EPFL Research and Teaching Assistant	<i>2011-now</i>
	Universitat Politècnica de Catalunya, UPC Teaching Assistant	<i>Jan'11-Aug'11</i>
	Universitat Nacional de Educación a Distancia, UNED Teaching Assistant	<i>Sept'10-June'11</i>
	GUIU Private College Teacher	<i>Sept'10-June'11</i>
	Caixa Penedès Financial Analyst	<i>Sept'08-Aug'10</i>
	Aula Escola Europea, IB World School Teacher	<i>Sept'07-Aug'8</i>
LANGUAGES	Catalan and Spanish (Mother Language), English (Fluent C1), French (Diplôme Approfondi en langue française, DALF C1)	
TECHNICAL SKILLS	Software: Matlab, Maple, Latex OS: Microsoft Windows, Mac, Linux	